GenCore version 5.1.3

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 14, 2003, 17:08:37 ; Search time 10 Seconds
(without alignments)
5393.494 Million cell updates/sec

Title: SEQ12-SEQ4
Perfect score: 2532
Sequence: 1 atgggattttttctcttttc.....gctgggcgcggcttaagctt 1390

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 237948

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/KUBELIK693/runat_14012003_091820_7146/app_query.fasta_1.1543
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=15 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=KUBELIK693_@CGN_1_1_6_@runat_14012003_091820_7146
-NCPU=6 -TCPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2079	82.1	403	10	US-09-086-118-23	Sequence 23, Appl
2	2079	82.1	403	10	US-09-835-684-3	Sequence 3, Appli
3	2079	82.1	403	10	US-09-880-371-3	Sequence 3, Appli
4	2079	82.1	403	10	US-09-879-248-3	Sequence 3, Appli
5	2079	82.1	403	10	US-09-770-693-3	Sequence 3, Appli
6	2079	82.1	403	10	US-09-766-348-3	Sequence 3, Appli
7	718.5	28.4	338	10	US-09-086-118-21	Sequence 21, Appl
8	718.5	28.4	338	10	US-09-835-684-1	Sequence 1, Appli
9	718.5	28.4	338	10	US-09-880-371-1	Sequence 1, Appli
10	718.5	28.4	338	10	US-09-879-248-1	Sequence 1, Appli
11	718.5	28.4	338	10	US-09-770-693-1	Sequence 1, Appli
12	718.5	28.4	338	10	US-09-766-348-1	Sequence 1, Appli
c 13	234.5	9.0	802	10	US-09-823-240-2	Sequence 2, Appli
c 14	220	8.5	731	9	US-10-086-464-17	Sequence 17, Appl
15	213.5	8.4	651	10	US-09-861-597-1	Sequence 1, Appli
c 16	212	8.2	1567	10	US-09-835-232-2	Sequence 2, Appli
17	205.5	8.1	606	10	US-09-861-597-6	Sequence 6, Appli
18	204.5	8.1	606	10	US-09-861-597-8	Sequence 8, Appli
c 19	202	7.8	280	10	US-09-835-232-1	Sequence 1, Appli
20	197.5	7.8	344	10	US-09-086-118-27	Sequence 27, Appl
21	197.5	7.8	344	10	US-09-835-684-11	Sequence 11, Appl
22	197.5	7.8	344	10	US-09-880-371-11	Sequence 11, Appl
23	197.5	7.8	344	10	US-09-879-248-15	Sequence 15, Appl
24	197.5	7.8	344	10	US-09-770-693-7	Sequence 7, Appli
25	197.5	7.8	344	10	US-09-766-348-7	Sequence 7, Appli
26	197.5	7.8	357	10	US-09-864-761-35807	Sequence 35807, A
27	197	7.8	606	10	US-09-861-597-4	Sequence 4, Appli
28	192.5	7.6	283	10	US-09-864-761-36720	Sequence 36720, A
c 29	183	7.1	647	10	US-09-925-299-1002	Sequence 1002, Ap
c 30	176.5	6.8	559	10	US-09-858-155A-2	Sequence 2, Appli
c 31	176	6.8	731	9	US-10-086-464-8	Sequence 8, Appli
32	174.5	6.9	440	9	US-10-066-500-106	Sequence 106, App
33	174.5	6.9	440	9	US-10-063-547-52	Sequence 52, Appl
34	174.5	6.9	440	12	US-10-006-867-52	Sequence 52, Appl
35	174.5	6.9	440	12	US-10-052-586-202	Sequence 202, App
c 36	174.5	6.7	5179	9	US-10-025-380-1068	Sequence 1068, Ap
c 37	174.5	6.7	5179	10	US-09-922-217-1068	Sequence 1068, Ap
c 38	174.5	6.7	5179	10	US-09-833-263-1068	Sequence 1068, Ap
c 39	172.5	6.7	503	9	US-10-078-547-2	Sequence 2, Appl
40	171	6.8	579	9	US-10-108-605-215	Sequence 215, App
41	169	6.7	34	10	US-09-770-693-13	Sequence 13, Appl
c 42	168	6.5	507	9	US-10-078-547-24	Sequence 24, Appl
43	166.5	6.6	191	10	US-09-864-761-36985	Sequence 36985, A
c 44	166	6.4	116	10	US-09-864-761-40290	Sequence 40290, A
45	165	6.5	34	10	US-09-770-693-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
US-09-086-118-23
; Sequence 23, Application US/09086118
; Patent No. US20010011380A1
;  GENERAL INFORMATION:
;    APPLICANT:  Laby, Ronald J.
;    APPLICANT:  Beer, Steven V.
;    APPLICANT:  Wei, Zhong-Min
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR
;    TITLE OF INVENTION:  FRAGMENTS ELICITING A HYPERSENSITIVE RESPONSE AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/086,118
;      FILING DATE:
;      CLASSIFICATION:
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/048,109
;      FILING DATE:  30-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1301
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-086-118-23

Alignment Scores:
Pred. No.:                1.02e-145     Length:          403
Score:                    2079.00       Matches:         403
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              82.11%        Indels:          0
DB:                       10            Gaps:            0

SEQ12-SEQ4 (1-1390) x US-09-086-118-23 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy    295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy    355 ACCGGCATGATGATGATGAGCATGATGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyLeuSerGlyGlyLeuGlyLeuGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGCACCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGGTACCGGTATCGGTIATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGCGATGGCGAAGCAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGCTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db    401 GlyAlaAla 403

RESULT 2
US-09-835-684-3
; Sequence 3, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Qiu, Dewen
;  APPLICANT:  Remick, Dean
;  TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
;  TITLE OF INVENTION:  RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
;  TITLE OF INVENTION:  DESICCATION
;  FILE REFERENCE: 21829/71
;  CURRENT APPLICATION NUMBER: US/09/835,684
;  CURRENT FILING DATE:  2001-04-16
;  PRIOR APPLICATION NUMBER: 60/198,359
;  PRIOR FILING DATE:  2000-04-19
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 403
;   TYPE: PRT
;   ORGANISM: Erwinia amylovora
US-09-835-684-3

Alignment Scores:
Pred. No.:                1.02e-145     Length:          403
Score:                    2079.00       Matches:         403
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              82.11%        Indels:          0
DB:                       10            Gaps:            0

SEQ12-SEQ4 (1-1390) x US-09-835-684-3 (1-403)
```

```
Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy    295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy    355 ACCGGCATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerSerSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCCGTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeuLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGACGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGGTACCGGTATCGGTATGGAAAGCGGGCATTCAGGCGCTCAATGATACGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGCCGATCGGGCGATGGCGAAGGAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGlnGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGGCAACGGCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db    401 GlyAlaAla 403

RESULT 3
US-09-880-371-3
; Sequence 3, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  DeRocher, Jay
;  TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE: 21829/91
;  CURRENT APPLICATION NUMBER: US/09/880,371
;  CURRENT FILING DATE:  2001-06-13
;  PRIOR APPLICATION NUMBER: 60/211,585
;  PRIOR FILING DATE: 2000-06-15
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;  LENGTH: 403
;  TYPE: PRT
;  ORGANISM: Erwinia amylovora
US-09-880-371-3

Alignment Scores:
Pred. No.:                  1.02e-145       Length:             403
Score:                      2079.00         Matches:            403
Percent Similarity:         100.00%         Conservative:         0
Best Local Similarity:      100.00%         Mismatches:           0
Query Match:                82.11%          Indels:               0
DB:                         10              Gaps:                 0

SEQ12-SEQ4 (1-1390) x US-09-880-371-3 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy    295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy    355 ACCGGCATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy     655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy     715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy     775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy     835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy     895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy     955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy    1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy    1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy    1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy    1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy    1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy    1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy    1375 GGCGCGGCT 1383
           ||||||||
Db     401 GlyAlaAla 403
```

RESULT 4
US-09-879-248-3
; Sequence 3, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 403
;   TYPE: PRT
;   ORGANISM: Erwinia amylovora
US-09-879-248-3

```
Alignment Scores:
Pred. No.:                1.02e-145     Length:         403
Score:                    2079.00       Matches:        403
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              82.11%        Indels:         0
DB:                       10            Gaps:           0
```

SEQ12-SEQ4 (1-1390) x US-09-879-248-3 (1-403)

```
Qy     175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy     235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy     295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy     355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy     415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy     475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy     535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy     595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy     655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy     715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy     775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy     835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy     895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy     955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy    1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300
```

```
Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db    401 GlyAlaAla 403

RESULT 5
US-09-770-693-3
; Sequence 3, Application US/09770693
; Patent No. US20020069434A1
; GENERAL INFORMATION:
;  APPLICANT: Beer, Steven V.
;  APPLICANT:  Bauer, David W.
;  TITLE OF INVENTION: OOMYCETE-RESISTANT TRANSGENIC PLANTS BY VIRTUE OF
;  TITLE OF INVENTION:  PATHOGEN-INDUCED EXPRESSION OF A HETEROLOGOUS
;  TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR
;  FILE REFERENCE: 19603/2501
;  CURRENT APPLICATION NUMBER: US/09/770,693
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: 60/178,565
;  PRIOR FILING DATE: 2000-01-26
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;    LENGTH: 403
;    TYPE: PRT
;    ORGANISM: Erwinia amylovora
US-09-770-693-3

Alignment Scores:
Pred. No.:                1.02e-145       Length:                 403
Score:                    2079.00         Matches:                403
Percent Similarity:       100.00%         Conservative:           0
Best Local Similarity:    100.00%         Mismatches:             0
Query Match:              82.11%          Indels:                 0
DB:                       10              Gaps:                   0

SEQ12-SEQ4 (1-1390) x US-09-770-693-3 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy    355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db    401 GlyAlaAla 403

RESULT 6
US-09-766-348-3
; Sequence 3, Application US/09766348
; Patent No. US20020116733A1
```

```
; GENERAL INFORMATION:
;  APPLICANT:  Qiu, Dewen
;  APPLICANT:  Wei, Zhong-Min
;  APPLICANT:  Beer, Steven V.
;  TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE INDUCED RESISTANCE IN PLANTS BY
;  TITLE OF INVENTION:  SEED TREATMENT
;  FILE REFERENCE:  19603/2986
;  CURRENT APPLICATION NUMBER: US/09/766,348
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER:  08/984,207
;  PRIOR FILING DATE:  1997-12-03
;  PRIOR APPLICATION NUMBER:  60/033,230
;  PRIOR FILING DATE:  1996-12-05
;  NUMBER OF SEQ ID NOS:  10
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 403
;   TYPE: PRT
;   ORGANISM: Erwinia amylovora
US-09-766-348-3

Alignment Scores:
Pred. No.:              1.02e-145       Length:          403
Score:                  2079.00         Matches:         403
Percent Similarity:     100.00%         Conservative:    0
Best Local Similarity:  100.00%         Mismatches:      0
Query Match:            82.11%          Indels:          0
DB:                     10              Gaps:            0

SEQ12-SEQ4 (1-1390) x US-09-766-348-3 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCCAAATTTCTATCGGCCGTCCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGGCTGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy    355 ACCGGCANGATGATGATGANGAGCATGATGGCGCTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyLeuMetGlyGlyLeu 80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTCAACACGCTGGGCTCGAAAGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGACGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuGlyAsn 220

Qy    835 GGCGGACTGGGAGGTGGTCAGCGCGGTAATCCTGCCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTCGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGGTACCGGTATCGGTATGAAACCGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCAATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGCCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGCTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATGGCCGCTGATGCCATTAACAATATCGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GCGCCGCCT 1383
          |||||||||
Db    401 GlyAlaAla 403

RESULT 7
US-09-086-118-21
; Sequence 21, Application US/09086118
; Patent No. US20010013804A1
;  GENERAL INFORMATION:
;    APPLICANT:  Laby, Ronald J.
;    APPLICANT:  Beer, Steven V.
;    APPLICANT:  Wei, Zhong-Min
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR
;    TITLE OF INVENTION:  FRAGMENTS ELICITING A HYPERSENSITIVE RESPONSE AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/086,118
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/048,109
;      FILING DATE:  30-MAY-1997
```

```
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Goldman, Michael L.
;       REGISTRATION NUMBER:  30,727
;       REFERENCE/DOCKET NUMBER:  19603/1301
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 263-1304
;       TELEFAX:  (716) 263-1600
;   INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  338 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-086-118-21

Alignment Scores:
Pred. No.:               1.29e-45      Length:         338
Score:                   718.50        Matches:        173
Percent Similarity:      52.97%        Conservative:   41
Best Local Similarity:   42.82%        Mismatches:     111
Query Match:             28.38%        Indels:         79
DB:                      10            Gaps:           11

SEQ12-SEQ4 (1-1390) x US-09-086-118-21 (1-338)

Qy    211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
          |||||||||:::|||                |||||:::   |||:::    |||
Db      1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy    268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
          |||      ||||||   |||   |||||||||          ||||||
Db     21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy    316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
             :::   |||:::::::::|||          ||||||   ||||||
Db     39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy    376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                  ||||||   |||       |||
Db     56 -----------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy    436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
          ||||||   |||               |||   |||:::||||||   |||         |||
Db     64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy    496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                                |||:::::::|||   :::|||        |||
Db     83 Asn--------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy    556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                 ||||||
Db     97 -----------------------------------------------------SerGly 98

Qy    616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
             |||:::   |||                                 |||||||||   :::
Db     99 GlyAspAlaLeuSer-----------------------------LysMetPheAspLys 108

Qy    676 ATAATGCAAAGCCTGTTTGGT------------------------GATGGGCAAGATGGC 711
             :::        |||   |||                           :::     :::
Db    109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy    712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                   |||                |||   |||:::|||       ||||||:::
Db    129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy    772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
             ||||||   :::|||||||||   ::::::||||||||||||   |||      :::
Db    149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy    832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                     :::|||          |||
Db    168 ---------------------------------------SerGlyPheSerGlnProSer 174

Qy    892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
          ||||||      |||||||||   |||||||||              :::   |||||||||||
Db    175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy    952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
          |||:::|||   |||:::|||        |||   :::   |||||:::::::::   |||
Db    195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy   1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
          |||         :::   |||   ||||||:::|||   ||||||   |||||||||||
Db    215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy   1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
          ||||||||||||||||||||||||||||||:::||||||||||||:::|||||||||
Db    235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy   1132 CCGGGTCAGGAGGTGAAAACCGATGACAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                         ||||||||||||||||||||||||||||||||||||||||||
Db    255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy   1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
          |||||||||||||||   ||||||||||:::::::|||   :::|||   ||||||||||||
Db    275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy   1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
             :::||||||||||||||||||      ||||||      ||||||||||||||||:::
Db    295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy   1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
          ||||||||||||||||||   :::   ||||||   |||   ||||||:::|||||||||
Db    315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy   1372 CTGGGCGCGGCT 1383
          |||      |||
Db    335 LeuAlaAsnAla 338
```

```
RESULT 8
US-09-835-684-1
; Sequence 1, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
;  APPLICANT: Wei, Zhong-Min
;  APPLICANT:  Qiu, Dewen
;  APPLICANT:  Remick, Dean
;  TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
;  TITLE OF INVENTION:  RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
;  TITLE OF INVENTION:  DESICCATION
;  FILE REFERENCE: 21829/71
;  CURRENT APPLICATION NUMBER: US/09/835,684
;  CURRENT FILING DATE:  2001-04-16
;  PRIOR APPLICATION NUMBER: 60/198,359
;  PRIOR FILING DATE: 2000-04-19
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Erwinia chrysanthemi
US-09-835-684-1

Alignment Scores:
Pred. No.:               1.29e-45      Length:         338
Score:                   718.50        Matches:        173
Percent Similarity:      52.97%        Conservative:   41
Best Local Similarity:   42.82%        Mismatches:     111
Query Match:             28.38%        Indels:         79
DB:                      10            Gaps:           11

SEQ12-SEQ4 (1-1390) x US-09-835-684-1 (1-338)

Qy    211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
          |||||||||:::|||                |||||:::   |||:::    |||
```

```
Db        1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy      268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
            |||     ||||||   |||   ||||||||           ||||||
Db       21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy      316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                  :::    |||:::::::::|||          ||||||   ||||||
Db       39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy      376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                   ||||||   |||      |||
Db       56 ----------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy      436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
            ||||||   |||            |||   |||:::||||||   |||          |||
Db       64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy      496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                                  |||::::::|||   :::|||      |||
Db       83 Asn---------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy      556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                  ||||||
Db       97 ------------------------------------------------------SerGly 98

Qy      616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
               |||:::   |||                                  |||||||||   :::
Db       99 GlyAspAlaLeuSer-----------------------------LysMetPheAspLys 108

Qy      676 ATAATGCAAAGCCTGTTTGGT-----------------------GATGGGCAAGATGGC 711
               :::       |||   |||                         :::       :::
Db      109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy      712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                     |||                |||   |||:::|||      ||||||:::
Db      129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy      772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
               ||||||   :::|||||||||   ::::::||||||||||||   |||      :::
Db      149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy      832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                        :::|||         |||
Db      168 ------------------------------------SerGlyPheSerGlnProSer 174

Qy      892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
            ||||||   |||||||||   |||||||||         :::   ||||||||||||
Db      175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy      952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
            |||:::|||   |||:::|||         |||   :::   ||||||:::::::::   |||
Db      195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy     1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
            |||         :::   |||   |||||||:::|||   ||||||   ||||||||||||
Db      215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy     1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
            |||||||||||||||||||||||||||||||||:::|||||||||:::|||||||||
Db      235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy     1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                             ||||||||||||||||||||||||||||||||||||||||||||
Db      255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy     1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
            ||||||||||||||   |||||||||::::::|||   :::|||   ||||||||||||
Db      275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy     1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
               :::||||||||||||||||||   ||||||         |||||||||||||||:::
Db      295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy     1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
            ||||||||||||||||   :::   ||||||   |||   ||||||:::|||||||||
Db      315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy     1372 CTGGGCGCGGCT 1383
            |||      |||
Db      335 LeuAlaAsnAla 338
```

RESULT 9
US-09-880-371-1
; Sequence 1, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: DeRocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Erwinia chrysanthemi
US-09-880-371-1

```
Alignment Scores:
Pred. No.:                1.29e-45      Length:          338
Score:                    718.50        Matches:         173
Percent Similarity:       52.97%        Conservative:    41
Best Local Similarity:    42.82%        Mismatches:      111
Query Match:              28.38%        Indels:          79
DB:                       10            Gaps:            11

SEQ12-SEQ4 (1-1390) x US-09-880-371-1 (1-338)

Qy      211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
            ||||||||||:::|||      |||||||:::   |||:::   |||
Db        1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy      268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
            |||     ||||||   |||   ||||||||           ||||||
Db       21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy      316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                  :::    |||:::::::::|||          ||||||   ||||||
Db       39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy      376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                   ||||||   |||      |||
Db       56 ----------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy      436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
            ||||||   |||            |||   |||:::||||||   |||          |||
Db       64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy      496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                                  |||::::::|||   :::|||      |||
Db       83 Asn---------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy      556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                  ||||||
Db       97 ------------------------------------------------------SerGly 98

Qy      616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
               |||:::   |||                                  |||||||||   :::
Db       99 GlyAspAlaLeuSer-----------------------------LysMetPheAspLys 108
```

```
Qy    676 ATAATGCAAAGCCTGTTTGGT-----------------------GATGGGCAAGATGGC 711
              :::       |||   |||                      :::      :::
Db    109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy    712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                   |||           |||   |||:::|||        ||||||:::
Db    129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy    772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
              ||||||   :::|||||||||   ::::::|||||||||||      |||     :::
Db    149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy    832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                  :::|||              |||
Db    168 ------------------------------------SerGlyPheSerGlnProSer 174

Qy    892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
          ||||||      |||||||||   |||||||||              :::   |||||||||||
Db    175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy    952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
          |||:::|||   |||:::|||         |||   :::   ||||||:::::::::   |||
Db    195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy   1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
          |||         :::    |||   ||||||:::|||   ||||||   |||||||||||
Db    215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy   1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
          ||||||||||||||||||||||||||||||:::||||||||||||:::|||||||||
Db    235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy   1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                           |||||||||||||||||||||||||||||||||||||||||||||
Db    255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy   1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
          |||||||||||||||      |||||||||::::::|||     :::|||   ||||||||||||
Db    275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy   1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
                :::||||||||||||||||||     ||||||       ||||||||||||||||:::
Db    295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy   1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
          |||||||||||||||||     :::      ||||||     |||     ||||||:::|||||||||
Db    315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy   1372 CTGGGCGCGGCT 1383
          |||       |||
Db    335 LeuAlaAsnAla 338
```

RESULT 10
US-09-879-248-1
; Sequence 1, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Erwinia chrysanthemi
US-09-879-248-1

```
Alignment Scores:
Pred. No.:              1.29e-45     Length:        338
Score:                  718.50       Matches:       173
Percent Similarity:     52.97%       Conservative:  41
Best Local Similarity:  42.82%       Mismatches:    111
Query Match:            28.38%       Indels:        79
DB:                     10           Gaps:          11

SEQ12-SEQ4 (1-1390) x US-09-879-248-1 (1-338)

Qy    211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
          |||||||||:::|||                     |||||||:::   |||:::   |||
Db      1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy    268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
          |||      ||||||   |||   |||||||||                ||||||
Db     21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy    316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
              :::    |||:::::::::|||            ||||||   ||||||
Db     39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy    376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                    ||||||   |||       |||
Db     56 ------------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy    436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
          ||||||   |||              |||   |||:::||||||   |||            |||
Db     64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy    496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                                |||::::::|||   :::|||      |||
Db     83 Asn---------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy    556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                  ||||||
Db     97 ---------------------------------------------------SerGly 98

Qy    616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
          |||:::   |||                                    |||||||||   :::
Db     99 GlyAspAlaLeuSer---------------------------LysMetPheAspLys 108

Qy    676 ATAATGCAAAGCCTGTTTGGT-----------------------GATGGGCAAGATGGC 711
              :::       |||   |||                      :::      :::
Db    109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy    712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                   |||           |||   |||:::|||        ||||||:::
Db    129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy    772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
              ||||||   :::|||||||||   ::::::|||||||||||      |||     :::
Db    149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy    832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                  :::|||              |||
Db    168 ------------------------------------SerGlyPheSerGlnProSer 174

Qy    892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
          ||||||      |||||||||   |||||||||              :::   |||||||||||
Db    175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy    952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
          |||:::|||   |||:::|||         |||   :::   ||||||:::::::::   |||
Db    195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy   1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
          |||         :::    |||   ||||||:::|||   ||||||   |||||||||||
Db    215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy   1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
          ||||||||||||||||||||||||||||||:::||||||||||||:::|||||||||
```

```
Db     235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy    1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                          |||||||||||||||||||||||||||||||||||||||||||
Db     255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy    1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
           ||||||||||||||    ||||||||:::::|||    :::|||    |||||||||||
Db     275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy    1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
              :::||||||||||||||||   ||||||        |||||||||||||||:::
Db     295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy    1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
           |||||||||||||||||    :::    ||||||    |||    ||||||:::||||||||
Db     315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy    1372 CTGGGCGCGGCT 1383
             |||      |||
Db     335 LeuAlaAsnAla 338

RESULT 11
US-09-770-693-1
; Sequence 1, Application US/09770693
; Patent No. US20020069434A1
; GENERAL INFORMATION:
;  APPLICANT: Beer, Steven V.
;  APPLICANT:  Bauer, David W.
;  TITLE OF INVENTION: OOMYCETE-RESISTANT TRANSGENIC PLANTS BY VIRTUE OF
;  TITLE OF INVENTION:  PATHOGEN-INDUCED EXPRESSION OF A HETEROLOGOUS
;  TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR
;  FILE REFERENCE: 19603/2501
;  CURRENT APPLICATION NUMBER: US/09/770,693
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: 60/178,565
;  PRIOR FILING DATE: 2000-01-26
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Erwinia chrysanthemi
US-09-770-693-1

Alignment Scores:
Pred. No.:                1.29e-45      Length:          338
Score:                    718.50        Matches:         173
Percent Similarity:       52.97%        Conservative:    41
Best Local Similarity:    42.82%        Mismatches:      111
Query Match:              28.38%        Indels:          79
DB:                       10            Gaps:            11

SEQ12-SEQ4 (1-1390) x US-09-770-693-1 (1-338)

Qy     211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
           |||||||||:::|||               |||||||:::    |||:::    |||
Db       1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy     268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
            |||      ||||||    |||     |||||||||          ||||||
Db      21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy     316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                 :::    |||::::::::|||            ||||||    ||||||
Db      39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy     376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                   ||||||    |||        |||
Db      56 ----------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy     436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
           ||||||    |||              |||    |||:::||||||    |||          |||

Db      64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy     496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                                 |||::::::|||    :::|||         |||
Db      83 Asn---------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy     556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                  ||||||
Db      97 ----------------------------------------------------SerGly 98

Qy     616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
              |||:::   |||                              |||||||||   :::
Db      99 GlyAspAlaLeuSer-----------------------------LysMetPheAspLys 108

Qy     676 ATAATGCAAAGCCTGTTTGGT-----------------------GATGGGCAAGATGGC 711
              :::       |||   |||                          :::      :::
Db     109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy     712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                           |||      |||    |||:::|||          ||||||:::
Db     129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy     772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
              ||||||    :::|||||||||    ::::::||||||||||||   |||    :::
Db     149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy     832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                     :::|||          |||
Db     168 ---------------------------------------SerGlyPheSerGlnProSer 174

Qy     892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
           ||||||      |||||||||    |||||||||          :::    ||||||||||||
Db     175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy     952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
           |||:::|||   |||:::|||          |||    :::    |||||:::::::::   |||
Db     195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy    1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
           |||        :::    |||    ||||||:::|||    ||||||    |||||||||||
Db     215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy    1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
           ||||||||||||||||||||||||||||||||||||:::||||||||||:::||||||||
Db     235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy    1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                          |||||||||||||||||||||||||||||||||||||||||||
Db     255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy    1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
           ||||||||||||||    ||||||||:::::|||    :::|||    |||||||||||
Db     275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy    1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
              :::||||||||||||||||   ||||||        |||||||||||||||:::
Db     295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy    1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
           |||||||||||||||||    :::    ||||||    |||    ||||||:::||||||||
Db     315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy    1372 CTGGGCGCGGCT 1383
             |||      |||
Db     335 LeuAlaAsnAla 338

RESULT 12
US-09-766-348-1
; Sequence 1, Application US/09766348
; Patent No. US20020116733A1
; GENERAL INFORMATION:
;  APPLICANT:  Qiu, Dewen
;  APPLICANT:  Wei, Zhong-Min
```

```
;  APPLICANT:  Beer, Steven V.
;  TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE INDUCED RESISTANCE IN PLANTS BY
;  TITLE OF INVENTION:  SEED TREATMENT
;  FILE REFERENCE:  19603/2986
;  CURRENT APPLICATION NUMBER: US/09/766,348
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER:  08/984,207
;  PRIOR FILING DATE:  1997-12-03
;  PRIOR APPLICATION NUMBER:  60/033,230
;  PRIOR FILING DATE:  1996-12-05
;  NUMBER OF SEQ ID NOS:  10
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Erwinia chrysanthemi
US-09-766-348-1

Alignment Scores:
Pred. No.:              1.29e-45      Length:         338
Score:                  718.50        Matches:        173
Percent Similarity:     52.97%        Conservative:   41
Best Local Similarity:  42.82%        Mismatches:     111
Query Match:            28.38%        Indels:         79
DB:                     10            Gaps:           11

SEQ12-SEQ4 (1-1390) x US-09-766-348-1 (1-338)

Qy    211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
          |||||||||:::|||            ||||||:::   |||:::   |||
Db      1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy    268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
          |||      ||||||   |||   |||||||||          ||||||
Db     21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy    316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                :::   |||::::::::|||        ||||||   ||||||
Db     39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy    376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                 ||||||   |||       |||
Db     56 -----------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy    436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
          ||||||   |||         |||   |||:::||||||   |||            |||
Db     64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy    496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                            |||::::::|||   :::|||      |||
Db     83 Asn---------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy    556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                ||||||
Db     97 -----------------------------------------------------SerGly 98

Qy    616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
             |||:::   |||                                |||||||||   :::
Db     99 GlyAspAlaLeuSer-----------------------------LysMetPheAspLys 108

Qy    676 ATAATGCAAAGCCTGTTTGGT-----------------------GATGGCAAGATGGC 711
             :::      |||   |||                        :::         :::
Db    109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy    712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                   |||                  |||   |||:::|||        ||||||:::
Db    129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy    772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
             ||||||   :::|||||||||   ::::::||||||||||||   |||      :::
Db    149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy    832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                       :::|||           |||
Db    168 ---------------------------------------SerGlyPheSerGlnProSer 174

Qy    892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
          ||||||         |||||||||      |||||||||           :::   |||||||||||
Db    175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy    952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
          |||:::|||   |||:::|||         |||   :::   ||||||:::::::::   |||
Db    195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy   1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
          |||      :::   |||   ||||||:::|||   ||||||   |||||||||||
Db    215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy   1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
          |||||||||||||||||||||||||||:::||||||||||||:::|||||||||
Db    235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy   1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                         ||||||||||||||||||||||||||||||||||||||||||||||
Db    255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy   1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
          |||||||||||||||   |||||||||::::::|||   :::|||   ||||||||||||
Db    275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy   1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
             :::||||||||||||||||||   ||||||      |||||||||||||||:::
Db    295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy   1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
          |||||||||||||||||   :::   ||||||   |||   ||||||:::|||||||||
Db    315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy   1372 CTGGGCGCGGCT 1383
          |||      |||
Db    335 LeuAlaAsnAla 338

RESULT 13
US-09-823-240-2
; Sequence 2, Application US/09823240
; Patent No. US20020048813A1
; GENERAL INFORMATION:
;  APPLICANT: Frank B. Gertler
;  APPLICANT:  James E.Bear
;  APPLICANT:  Jurgen Wehland
;  APPLICANT:  Joseph Loureio
;  TITLE OF INVENTION: Methods and Products for Regulating Cell
;  TITLE OF INVENTION:  Motility
;  FILE REFERENCE: M0656/7064 (HCL)
;  CURRENT APPLICATION NUMBER: US/09/823,240
;  CURRENT FILING DATE:  2001-03-30
;  PRIOR APPLICATION NUMBER: 60/194,564
;  PRIOR FILING DATE: 2000-04-03
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 802
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-823-240-2

Alignment Scores:
Pred. No.:              6.25e-10      Length:         802
Score:                  234.50        Matches:        113
Percent Similarity:     34.20%        Conservative:   32
Best Local Similarity:  26.65%        Mismatches:     166
Query Match:            9.05%         Indels:         113
DB:                     10            Gaps:           20

SEQ12-SEQ4 (1-1390) x US-09-823-240-2 (1-802)
```

```
Qy   1382 GCCGCGCCCAGCTTGCCAAGTGCCATATTGTTAATGGCATCACCGGCCATCATGGCATCA 1323
          |||||||||||        |||:::::                |||        :::
Db    257 AlaAlaProSerSerAspSerSerLeuSerSerAlaProLeuProGluTyrSerSerCys 276

Qy   1322 ATACCCAGCGAAGAACCACCGGCACCGCGTGCCTGCAGGTTGCCGTTGCCGGTATCACCC 1263
             |||        ||||||      |||       :::   |||||||||
Db    277 GlnProProSerAlaProProProSerTyrAlaLysValIleSerAlaProValSerAsp 296

Qy   1262 GCCATGGGCCTTTTGATCATGCCCTTGGCTTTGTTGAACTGCTCCATACTGGCTGGTGTC 1203
          |||                                      :::   :::::::::      :::
Db    297 AlaThrPro-------------------------AspTyrAlaValValThrAlaLeu 307

Qy   1202 ATTCCGTCGTCATCTGGCTTGCTCAGTGCTTTTGCCCATGATTTGTCATCGGTTTTCACC 1143
             |||:::|||:::                   |||      :::        |||
Db    308 ProProThrSerThrProProThrProProLeuArgHisAlaAlaThrArgPheAlaThr 327

Qy   1142 TCCTGACCCGGGCCTTTCTGGTACTGCGGCTTGCCAAACACCTCAGGATACTGGTCCATG 1083
          |||      |||      ::::::      ||||||:::
Db    328 Ser---LeuGlySerAlaPheHisProValLeuProHis-------------------- 339

Qy   1082 AACTGACCGATTTCCTTCGCCATCGCCCGATCGCCTTTATTGACGAAAGAACGGGTTGAA 1023
                         :::|||           ||||||          |||
Db    340 --------------TyrAlaThrValProArgProLeuAsnLysAsnSerArg------ 352

Qy   1022 CTGTGCCTGTGCGTACCGATATCATTCAGCGCCTGAATGCCCGCTTTCATACCGATACCG 963
                         |||   |||              |||      |||
Db    353 --------------ProSerSer--------------ProValAsnThrProSerSer 362

Qy    962 GTACCCACGGCGTTACCTAACTGC---TGGTAGTCCACCGGCCCGCTCAGGTTTTGCAGC 906
             |||   |||      :::|||   |||           |||           |||
Db    363 GlnProProAlaAlaLysSerCysAlaTrp-----------ProThrSerAsnPheSer 378

Qy    905 CCTTTGCCGCCCAGC--------------------------GACGAACCGTCAAGACCC 873
          |||||||||||||||                           :::    :::   |||
Db    379 ProLeuProProSerProProIleMetIleSerSerProProGlyLysAlaThrGlyPro 398

Qy    872 GTGCCAGCATTACCG--------------------CCCTGACCACCTCCCAGTCCCCCG 834
             |||   ||||||                      |||      ||||||||||||
Db    399 ArgProValLeuProValCysValSerSerProValProGlnMetProProSerProThr 418

Qy    833 TTGCCAAGGAGCTGGCTC-----------AGACCATTACCCATCAGGCCCGACAGCGCA 786
             |||      |||                |||:::          |||   |||
Db    419 AlaProAsnGlySerLeuAspSerValThrTyrProValSerProProProThrSerGly 438

Qy    785 TCAGTGACTCCTTTTTATAGGCGTTCTGCTCGCCTTCGGTCGGCTGCTTGCCCCCAGAG 726
                |||                        |||             ||||||
Db    439 ProAlaAlaPro-----------------ProProProProProProProProProPro 452

Qy    725 GAACTGCCCTGGGTGCCATCTTGCCCATCACCAAACAGGCTTTGCATTATCTCGCTGAAC 666
                |||   :::|||     |||   |||                        |||
Db    453 ProProProProLeuProProProProLeuPro-----------------ProLeuAla 466

Qy    665 ATCTTCAGCAGCTGCTGCATCGGGTCGCTGGAGTCTGAGGTGGAATCTGTGCCGGAGGTG 606
                |||   |||      ||||||      |||               |||
Db    467 SerLeuSerHisCys------GlySerGlnAlaSerProProProGlyThrProLeuAla 484

Qy    605 GAATCGTCGTTTTGGGACGTTGAGTTAATACCCAGCGCCTGGTCCAGCGGGGAATTTGTT 546
             :::                 ::::::||||||        :::
Db    485 SerThrProSerSerLysProSerValLeuProSerProSerAlaGlyAlaProAlaSer 504

Qy    545 GTTGAAGTGGTATTGTTGCCGCCTTTCGAGCCCAGCGTGTTCAGCGAACCGCCT---AAC 489
             |||      |||   |||                 |||      ||||||||
Db    505 AlaGluThrProLeuAsnProGluLeuGlyAspSerSerAlaSerGluProGlyLeuGln 524

Qy    488 ATATCGTTCAGCGCGTTCGACAGTCCTTCGCCCAGG--------------------CCA 450
             :::            :::||||||:::|||:::
Db    525 AlaAlaSerGlnProAlaGluSerProThrProGlnGlyLeuValLeuGlyProProAla 544

Qy    449 CCTGAGCCACCCAAGCCATTACCTAAG-----------------------CCACCGCCT 414
          |||   ||||||   |||||||||                          |||||||||
Db    545 ProProProProProProLeuProSerGlyProAlaTyrAlaSerAlaLeuProProPro 564

Qy    413 AAGCCACCGCCCATCAGCCCACCACCGCCCATCATGCTCATCATCATCATCATGCCGGTG 354
             ||||||      ||||||||||||                            :::|||
Db    565 ProGlyProPro------ProProProPro--------------------LeuProSer 575

Qy    353 AGTAAGCCAGCCAGCTGATTGACGGTATCATTTTGATTACCGCCGCCCAGCCCCAGTGCA 294
          :::   |||                                   |||||||||   |||
Db    576 ThrGlyPro-----------------------------ProProProProProProPro 585

Qy    293 GAATTGCCACCCAACCCAGCATTCTGGCGACTGGTACCCAGCAACCCGTTATTTCCGCCC 234
                ||||||    |||                   |||       |||       ||||||
Db    586 ProProLeuProAsnGlnAla--------------ProProProProProProProPro 600

Qy    233 GCACCGCCGATA 222
          |||||||||:::
Db    601 AlaProProLeu 604
```

RESULT 14
US-10-086-464-17
; Sequence 17, Application US/10086464
; Publication No. US20020199218A1
; GENERAL INFORMATION:
; APPLICANT: GORING, Daphne R. et al.
; TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES
; FILE REFERENCE: P 25,762-A USA
; CURRENT APPLICATION NUMBER: US/10/086,464
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 10/069,304
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: PCT/CA00/00966
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/149,466
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: US 60/159,122
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
;   LENGTH: 731
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-086-464-17

```
Alignment Scores:
Pred. No.:                7.17e-09      Length:        731
Score:                    220.00        Matches:       97
Percent Similarity:       37.27%        Conservative:  23
Best Local Similarity:    30.12%        Mismatches:    132
Query Match:              8.49%         Indels:        70
DB:                       9             Gaps:          13
```

SEQ12-SEQ4 (1-1390) x US-10-086-464-17 (1-731)

```
Qy   1147 TCACCTCCTGACCCGGGCCTTTCTGGTACTGCGGCTTGCCAAACACCTCAGGATACTGGT 1088
          |||   |||   |||         |||         |||               |||   |||:::
Db      7 SerSerProAlaProAlaThrSerProProAlaMetSerLeuProProAlaAspSerVal 26

Qy   1087 CCATGAACTGACCGATTTCCTTCGCCATCGCCCGATCGCCTTTATTGACGAAAGAACGGG 1028
          |||   |||         |||:::|||         |||||||||
Db     27 ProAspThrSerSerProProAlaPro--ProLeuSerProLeuProPro---------- 42

Qy   1027 TTGAACTGTGCCTGTGCGTACCGATATCATTCAGCGCCTGAATGCCCGCTTTCATACCGA 968
                            |||:::|||          :::|||:::         |||:
Db     43 ------------------ProLeuSerSerProProProLeuProSerProProProL 56

Qy    967 TACCGGTACCCACGGCGTTACCTAACTGCTGGTAGTCCACCGGCCCGCTCAGGTTTTGCA 908
          ::       |||||||||   |||                        ||||||        |
Db     56 euSerAlaProThrAlaSerPro-----------------ProProLeuProValGluS 70

Qy    907 GC---CCTTTGCCGCCCAGCGACGAACCGTCAAGACCCGTGCCAGCATTACCGCCCTGAC 851
          ||     |||    ||||||     :::    |||      |||:::           ||||||   |
Db     70 erProProSerProProIleGluSerProProProProLeuLeuGluSerProProProP 90

Qy    850 CACCT---------CCCAGTCCCCCGTTGCCAAGGAGCTGGCTCAGACCATTACCCATCA 800
          |||||           ||||||||||||   |||
```

```
Db     90 roProLeuGluSerProSerProProSerProHis----------------ValSerA 104

Qy    799 GGCCCGACAGCGCATCAGTGACTCCTTTTTATAGGCGTTCTGCTCGCCTTCGGTCGGCT 740
Db    104 laProSerGlySerProProLeuProPheLeuProAlaLysProSerProProProSerS 124

Qy    739 GCTTGCCCCCAGAGGAACTGCCCTGGGTGCCATCTTGCCCATCACCAAACAGGCTTTGCA 680
Db    124 erProProSerGluThrValProProGlyAsnThrIleSerProProProArg------- 141

Qy    679 TTATCTCGCTCAATCTTCAGCAGCTGCTGCATCGAGGTCGCTGGAGTCTGAGGTGGAAT 620
Db    142 ----------------------------------------SerLeuProSer------GluS 147

Qy    619 CTGTGCCGGAGGTGGAATCGTCGTTTTGGGACGTTGAGTTAATACCCAGCGCCTGGTCCA 560
Db    147 erThrProProValAsnThrAlaSerProProProProSerProProArgArgArgSerG 167

Qy    559 GCGGGGAATTTGTTGTTGAAGTGGTATTGTTGCCGCCTTTCGGAGCCCAGCGTGTTCAGCG 500
Db    167 lyProLys----------------------ProSerPheProProProProIleAsnSerS 179

Qy    499 AACCGCCTAACATATCGTTCAGCGCGTTCGACAGTCCTTCGCCCAGGCCACCTGAGCCAC 440
Db    179 erProProAsn----------------------ProSerProAsnThrProSerLeuP 191

Qy    439 CCAAG------CCATTACCTAAGCCACCG-----------CCTAAGCCACCGCCCATCA 398
Db    191 roGluThrSerProProProLysProProLeuSerThrThrProPheProSerSerSerT 211

Qy    397 GCCCACCACCGCCCATCATGCTCATCATCATCATGCCGGTGAGTAAGCCAGCCAGCT 338
Db    211 hrProProProLysLysSerProAlaAlaValThrLeuProPhePheGlyProAlaGlyG 231

Qy    337 GATTGACGGTATCATTTTGATTACCGCCGCCCAGCCCCAGTGCAGAATTGCCACCCAACC 278
Db    231 lnLeuProAspGlyThrValAlaProProIleGlyProValIleGluProLysThrSerP 251

Qy    277 CAGCATTCTGGCAGCTGGTACCCAGC------AACCCGTTATTTCCGCCCGCACCGCCGA 224
Db    251 roAla---GluSerIleSerProGlyThrProGlnProLeuValProLysSerLeuProV 270

Qy    223 TA 222
Db    270 al 270
```

```
RESULT 15
US-09-861-597-1
; Sequence 1, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
;  APPLICANT: PHILLIPPE, Michel
;  APPLICANT:  GARSON, Jean-Claude
;  APPLICANT:  ARRAUDEAU, Jean-Pierre
;  TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
;  TITLE OF INVENTION:  LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
;  TITLE OF INVENTION:  ANALOG
;  FILE REFERENCE: 6388-0365-0
;  CURRENT APPLICATION NUMBER: US/09/861,597
;  CURRENT FILING DATE:  2001-05-22
;  PRIOR APPLICATION NUMBER: 09/247,806
;  PRIOR FILING DATE: 1999-02-11
;  PRIOR APPLICATION NUMBER: FR 98/01614
;  PRIOR FILING DATE: 1998-02-11
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;    LENGTH: 651
;    TYPE: PRT
;    ORGANISM: Nephila clavipes
US-09-861-597-1

Alignment Scores:
```

```
Pred. No.:              2.11e-08        Length:         651
Score:                  213.50          Matches:        118
Percent Similarity:     33.33%          Conservative:   20
Best Local Similarity:  28.50%          Mismatches:     196
Query Match:            8.43%           Indels:         81
DB:                     10              Gaps:           18

SEQ12-SEQ4 (1-1390) x US-09-861-597-1 (1-651)

Qy    193 GGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAATAACGGTTG 252
Db    104 GlyGlnGlyAlaSlyAlaAlaAlaAlaAlaAlaAlaGlyGlyAlaGlyGlnGlyGlyTyr 123

Qy    253 CTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTGCACTG------------ 300
Db    124 GlyGlyLeuGlyAsnGlnGlyAlaGlyArgGlyGlyGlnGlyAlaAlaAlaAlaAlaAla 143

Qy    301 ---GGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTA------ 351
Db    144 GlyGlyAlaGlyGlnGlyGlyTyrGlyGlyLeuGlySerGlnGlyAlaGlyArgGlyGly 163

Qy    352 CTCACCGGCATGATGATGATGATGGACGTGGTTGCGTGAACACGCTGGGCTCGAAAGGC 411
Db    164 LeuGlyGlyGlnGlyAlaGlyAlaAlaAlaAlaAlaAlaGlyGlyAlaGlyGlnGlyGly 183

Qy    412 TTAGGC------GGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGA 465
Db    184 TyrGlyGlyLeuGlyGlyGlnGlyAlaGlyGlnGlyGlyTyrGlyGlyLeuGlySer--- 202

Qy    466 CTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGC 525
Db    203 ----------------------GlnGlyAlaGlyArgGlyGlyLeuGlyGlyGlnGly 214

Qy    526 GGCAACAATACCACTTCAACAACAAATTCCCCGCTGGACCAG---GCGCCTGGGTATTAAC 582
Db    215 AlaGlyAlaAlaAlaAlaAlaAlaAlaGlyGlyAlaGlyGlnGlyGlyLeuGlyGlyGln 234

Qy    583 TCAACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGAC 642
Db    235 GlyAlaGlyGlnGlyAlaGlyAlaSerAlaAlaAlaAlaGlyGlyAlaGlyGlnGly--- 253

Qy    643 CCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGG 702
Db    254 ----------------------------------------------------GlyTyrGly 256

Qy    703 CAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAAC 762
Db    257 GlyLeuGlySerGlnGlyAlaGlyArgGlyGly----------GluGly---------- 269

Qy    763 GCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAG 822
Db    270 ---------AlaGlyAlaAlaAlaAlaAlaAlaGlyGlyAlaGlyGlnGly--------- 283

Qy    823 CTCCTTGGCAACGGGGGACTGGGAGGT---------GGTCAGGGCGGTAATGCTGGCACG 873
Db    284 -----GlyTyrGlyGlyLeuGlyGlyGlnGlyAlaGlyGlnGlyGlyTyrGlyGlyLeu 301

Qy    874 GGTCTTGACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAAC----------CTGAGC 921
Db    302 GlySerGlnGlyAlaGlyArgGlyGlyLeuGlyGlyGlnGlyAlaGlyAlaAlaAlaAla 321

Qy    922 GGGCCGGTGGACTACCAGCAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATAAAGCG 981
Db    322 GlyGlyAlaGlyGlnGlyGlyLeuGly---------GlyGlnGlyAlaGlyGlnGlyAla 338

Qy    982 GGCATTCAGCGCTGAATGATATCGGTAC-GCACAGGCACAGTTCAACCCGTTCTTTCGT 1040
Db    339 GlyAlaAlaAlaAlaAlaGlyGlyAlaGlyGlnGlyGlyTyrGlyGlyLeuGlySer 358

Qy   1041 CAATAAAGGCGATCGGGCGATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCTGA 1100
Db    359 Gln----------GlyAlaGlyArgGlyGlyLeuGlyGlyGlnGlyAla--------- 371
```

```
Qy    1101 GGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGACAA 1160
           |||          ||||||                 |||   ||||||
Db     372 GlyAlaValAlaAlaAlaAlaAlaGlyGlyAlaGlyGlnGlyGlyTyrGly--------- 388

Qy    1161 ATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATGGA 1220
             |||          ::::::   |||          |||
Db     389 ------GlyLeuGlySerGlnGlyAla------GlyArgGlyGlyGlnGlyAlaGlyAla 404

Qy    1221 GCAGTTCAACAAAGCCAAGGGCATGATCAAAAGGCCCATGGCGGGTGATACCGGCAACGG 1280
           |||          :::   |||          ||:::   ::::||||||
Db     405 AlaAlaAlaAlaAlaGlyGlyAlaGlyGlnArgGlyTyrGlyGly------LeuGlyAsn 422

Qy    1281 CAACCTGCAGGCACGCGGTGCC---GGTGGTTCTTCGCTGGGTATTGATGCCATGATGGC 1337
           |||  ||||| ||||||     ||||||          |||     |||     ||
Db     423 GlnGlyAlaGly-ArgGlyGlyLeuGlyGlyGlnGlyAlaGlyAlaAlaAlaAlaAl 442

Qy    1338 CGGTGATGCCATTAACAATATGGCACTTGGCAAGCTGGGC 1377
           |                          |||   ||||||
Db     442 aAlaGlyGlyAlaGlyGlnGlyGlyTyrGlyGlyLeuGly 455

Search completed: January 14, 2003, 17:18:53
Job time : 18 secs
```

GenCore version 5.1.3

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 14, 2003, 17:09:17 ; Search time 13.5 Seconds
(without alignments)
6058.942 Million cell updates/sec

Title: SEQ12-SEQ4
Perfect score: 2532
Sequence: 1 atgggatttttttctttttc.....gctgggcgcggcttaagctt 1390

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/KUBELIK693/runat_14012003_091819_7104/app_query.fasta_1.1543
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=KUBELIK693_SCGN_1_1_11_8runat_14012003_091819_7104 -NCPU=6 -1CPU=3
-NO_XLFXY -NO_MMAP -LASTQUERY -NSG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2079	82.1	403	2	US-08-200-724A-2	Sequence 2, Appli
2	2079	82.1	403	2	US-09-030-270A-3	Sequence 3, Appli
3	2079	82.1	403	4	US-08-851-376A-2	Sequence 2, Appli
4	2079	82.1	403	4	US-08-984-207-3	Sequence 3, Appli
5	2079	82.1	403	4	US-09-013-587-3	Sequence 3, Appli
6	1986	78.4	385	1	US-08-891-254-3	Sequence 3, Appli
7	1986	78.4	385	2	US-08-819-539-3	Sequence 3, Appli
8	1986	78.4	385	5	PCT-US96-08819-3	Sequence 3, Appli
9	1962	77.5	385	5	PCT-US93-06243-2	Sequence 2, Appli
10	718.5	28.4	338	1	US-08-891-254-1	Sequence 1, Appli
11	718.5	28.4	338	2	US-08-484-358-2	Sequence 2, Appli
12	718.5	28.4	338	2	US-08-819-539-1	Sequence 1, Appli
13	718.5	28.4	338	2	US-09-030-270A-1	Sequence 1, Appli
14	718.5	28.4	338	3	US-09-118-959-2	Sequence 2, Appli
15	718.5	28.4	338	4	US-08-984-207-1	Sequence 1, Appli
16	718.5	28.4	338	4	US-09-013-587-1	Sequence 1, Appli
17	718.5	28.4	338	5	PCT-US96-08819-1	Sequence 1, Appli
18	213.5	8.4	604	4	US-08-556-978B-63	Sequence 63, Appl
19	213.5	8.4	651	4	US-08-556-978B-19	Sequence 19, Appl
20	213.5	8.4	651	4	US-09-247-806-1	Sequence 1, Appli
21	213.5	8.4	718	1	US-08-425-069-2	Sequence 2, Appli
22	213.5	8.4	718	2	US-08-317-844B-2	Sequence 2, Appli
23	213.5	8.4	747	3	US-09-034-177-3	Sequence 3, Appli
24	211.5	8.4	318	4	US-09-060-756-727	Sequence 727, App
25	211	8.3	738	3	US-08-864-038A-3	Sequence 3, Appli
26	205.5	8.1	606	4	US-09-247-806-6	Sequence 6, Appli
27	204.5	8.1	606	4	US-08-556-978B-23	Sequence 23, Appl
28	204.5	8.1	606	4	US-09-247-806-8	Sequence 8, Appli
c 29	204	7.9	1255	2	US-09-080-897-4	Sequence 4, Appli
c 30	204	7.9	1255	3	US-08-899-595-1	Sequence 1, Appli
c 31	204	7.9	1255	4	US-09-323-735-4	Sequence 4, Appli
32	203	8.0	334	4	US-09-060-756-728	Sequence 728, App
33	203	8.0	641	4	US-09-249-585A-3	Sequence 3, Appli
c 34	201	7.8	1248	2	US-09-080-897-2	Sequence 2, Appli
c 35	201	7.8	1248	4	US-09-323-735-2	Sequence 2, Appli
c 36	201	7.8	1315	3	US-08-899-595-3	Sequence 3, Appli
c 37	200.5	7.7	174	4	US-08-818-112-143	Sequence 143, App
c 38	200.5	7.7	174	4	US-08-818-111-138	Sequence 138, App
c 39	200.5	7.7	174	4	US-09-056-556-143	Sequence 143, App
c 40	200.5	7.7	174	4	US-09-072-596-138	Sequence 138, App
41	197.5	7.8	344	1	US-08-891-254-7	Sequence 7, Appli
42	197.5	7.8	344	2	US-08-819-539-7	Sequence 7, Appli
43	197.5	7.8	344	2	US-09-030-270A-7	Sequence 7, Appli
44	197.5	7.8	344	4	US-08-984-207-7	Sequence 7, Appli
45	197.5	7.8	344	4	US-09-013-587-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
US-08-200-724A-2
; Sequence 2, Application US/08200724A
; Patent No. 5849868
;  GENERAL INFORMATION:
;    APPLICANT:  Wei, Zhong-Min
;    APPLICANT:  Bauer, David W.
;    APPLICANT:  Beer, Steven V.
;    APPLICANT:  Collmer, Alan
;    APPLICANT:  He, Sheng-Yang
;    APPLICANT:  Laby, Ron J.
;    TITLE OF INVENTION:  ELICITOR OF THE HYPERSENSITIVE RESPONSE
;    TITLE OF INVENTION:  IN PLANTS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/200,724A
;      FILING DATE:  23-FEB-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/10030
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
```

```
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      STRANEDENESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-200-724A-2

Alignment Scores:
Pred. No.:                5.54e-176       Length:          403
Score:                    2079.00         Matches:         403
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              82.11%          Indels:          0
DB:                       2               Gaps:            0

SEQ12-SEQ4 (1-1390) x US-08-200-724A-2 (1-403)

Qy     175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy     235 GGCGGAAATAACGGGTTGCTGGTACCAGTCGCCAGAATGCYGGGTTGGGTGGCAATTCT 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy     295 GCACTGGGGCTGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy     355 ACCGGCATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ThrGlyMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy     415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy     475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAAACAAT 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerGlyGlyAsnAsn  120

Qy     535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy     595 AACGACGATTCCACCTCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy     655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy     715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy     775 GGAGTCACTGATGCGCCTGTCGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTGGCAAC 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuGlyAsn  220

Qy     835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy     895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy     955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATCGGTACGCAC 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy    1015 AGGCACAGTTCAACCCGTTCTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy    1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy    1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy    1195 GACGGAATGACACCAGCCACTATGGACCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy    1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTCGCAGGCACGCGTGCCGGTGGTTCTTCC 1314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy    1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy    1375 GGCGCGGCT 1383
           |||||||||
Db     401 GlyAlaAla 403

RESULT 2
US-09-030-270A-3
; Sequence 3, Application US/09030270A
; Patent No. 5977060
;  GENERAL INFORMATION:
;    APPLICANT:  Zitter, Thomas A.
;    APPLICANT:  Wei, Zhong-Min
;    TITLE OF INVENTION:  INSECT CONTROL WITH A
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,270A
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/039,226
;      FILING DATE:  28-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1521
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      STRANEDENESS:
;      TOPOLOGY:  linear
```

```
;     MOLECULE TYPE:  protein
US-09-030-270A-3

Alignment Scores:
Pred. No.:              5.54e-176       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
DB:                     2               Gaps:           0

SEQ12-SEQ4 (1-1390) x US-09-030-270A-3 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy    295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy    355 ACCGGCATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGCGCTCGAAAGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCCAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGlnIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGCGCAGCACAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGACGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGGTACCGGTATCGGTATGAAAGCGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGCCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATGGGCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db     401 GlyAlaAla 403

RESULT 3
US-08-851-376A-2
; Sequence 2, Application US/08851376A
; Patent No. 6174717
;  GENERAL INFORMATION:
;    APPLICANT:  Beer, Steven V.
;    APPLICANT:  Wei, Zhong-Min
;    APPLICANT:  Bauer, David W.
;    APPLICANT:  Collmer, Alan
;    APPLICANT:  He, Sheng-Yang
;    APPLICANT:  Laby, Ron
;    TITLE OF INVENTION:  ELICITOR OF THE HYPERSENSITIVE RESPONSE
;    TITLE OF INVENTION:  IN PLANTS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon Peabody LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  NY
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/851,376A
;      FILING DATE:  05-MAY-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/200,724
;      FILING DATE:  23-FEB-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/10035
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-851-376A-2
```

```
Alignment Scores:
Pred. No.:              5.54e-176       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
DB:                     4               Gaps:           0

SEQ12-SEQ4 (1-1390) x US-08-851-376A-2 (1-403)

Qy     175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy     235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCCAGAATGCTGGGTTGGGTGGCAATTCT 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy     295 GCACTGGGGCTGGCGGCCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy     355 ACCGGCATGATGATGATGAGCATGATGGCCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ThrGlyMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy     415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTCTCGAAC 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy     475 GCGCTGAAGCATATGTTAGGCGGTTCGCTGAACACCCTGGGCTCGAAAGGCGGCAACAAT 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy     535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy     595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGCACCCGATGCAGCAG 654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy     655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGCTGATGGGCAACATGCCACC 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy     715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy     775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlySerGlnLeuLeuGlyAlaAsn 220

Qy     835 GGGCGACTCGGAGCTGGTCAGGGCGTAATGCTCGCACGGGTCTTGACGGTTCGTCCGCTG 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy     895 GGCGGCAAAGGCCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy     955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy    1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGCGGATCGGGCGATGGCGAAGGAAATC 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy    1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy    1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy    1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy    1255 CCCATGGCGGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy    1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy    1375 GGCGCGGCT 1383
           |||||||||
Db     401 GlyAlaAla 403

RESULT 4
US-08-984-207-3
; Sequence 3, Application US/08984207
; Patent No. 6235974
;  GENERAL INFORMATION:
;     APPLICANT:  Qiu, Dewen
;     APPLICANT:  Wei, Zhong-Min
;     APPLICANT:  Beer, Steven V.
;     TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE INDUCED
;     TITLE OF INVENTION:  RESISTANCE IN PLANTS BY SEED TREATMENT
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;        STREET:  P.O. Box 1051, Clinton Square
;        CITY:  Rochester
;        STATE:  New York
;        COUNTRY:  U.S.A.
;        ZIP:  14603
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE:  Floppy disk
;        COMPUTER:  IBM PC compatible
;        OPERATING SYSTEM:  PC-DOS/MS-DOS
;        SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER:  US/08/984,207
;        FILING DATE:
;        CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;        APPLICATION NUMBER:  US 60/033,230
;        FILING DATE:  05-DEC-1996
;     ATTORNEY/AGENT INFORMATION:
;        NAME:  Goldman, Michael L.
;        REGISTRATION NUMBER:  30,727
;        REFERENCE/DOCKET NUMBER:  19603/1201
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE:  (716) 263-1304
;        TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH:  403 amino acids
;        TYPE:  amino acid
;        STRANDEDNESS:
;        TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-984-207-3

Alignment Scores:
Pred. No.:              5.54e-176       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
```

```
DB:          4       Gaps:       0

SEQ12-SEQ4 (1-1390) x US-08-984-207-3 (1-403)

Qy   175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy   235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy   295 GCACTGGGGCTGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy   355 ACCGGCATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ThrGlyMetMetMetMetSerMetMetGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy   415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy   475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGCCGGCAACAAT 534
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 AlaLeuAsnAspMetLeuGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy   535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy   595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy   655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy   715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy   775 GGAGTCACTGATGCGCTGTCGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy   835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   221 GlyGlyLeuGlyGlyGlyGlnGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy   895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy   955 GTGGGTACCGGTATCGGTATGCAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy  1015 AGCCACAGTTCAACCCGTTCTTTCGTCAATAAAGCGATCGGGCGATGGCGAAGGAAATC 1074
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy  1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy  1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy  1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy  1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy  1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy  1375 GGCGCGGCT 1383
         |||||||||
Db   401 GlyAlaAla 403

RESULT 5
US-09-013-587-3
; Sequence 3, Application US/09013587
; Patent No. 6277814
;  GENERAL INFORMATION:
;    APPLICANT:  Qiu, Dewen
;    APPLICANT:  Wei, Zhong-Min
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  ENHANCEMENT OF GROWTH IN PLANTS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentln Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/013,587
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/036,048
;      FILING DATE:  27-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1501
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-013-587-3

Alignment Scores:
Pred. No.:              5.54e-176       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
DB:                     4               Gaps:           0

SEQ12-SEQ4 (1-1390) x US-09-013-587-3 (1-403)

Qy   175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20
```

```
Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCTAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGGCTGGGCGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTTGCCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy    355 ACCGGCATGATCATGATGATGACATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCCAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyLeuGlyGlySerGlyGlyLeuGlyGluLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTCAACACGCTGGCCTCGAAAGGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGCACCAGCGCGCTGGCTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCCAGATAATGCAAACCCTGTTTGCTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGlnIleMetGlnSerLeuPheGlyAspGlyAsnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACCGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGCGCTGCAAAACCTGAGCGGGCCCGGTGGACTACCAGCAGTTAGGTAACCCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnLeuGlyAsnAla 260

Qy    955 CTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATGGGCGATGGCGAAGGAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCATGCACCAGTATCCTGAGGTCTTTGGCAAGCCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGCCAACCTGCAGGCACGCGGTGCCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380
```

```
Qy   1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db    401 GlyAlaAla 403

RESULT 6
US-08-891-254-3
; Sequence 3, Application US/08891254
; Patent No. 5776889
;  GENERAL INFORMATION:
;     APPLICANT:  Wei, Zhong-Min
;     APPLICANT:  Beer, Steven V.
;     TITLE OF INVENTION:  Hypersensitive Response
;     TITLE OF INVENTION:  Induced Resistance In Plants
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE:  Wixon, Hargrave, Devans & Doyle
;        STREET:  Clinton Square, P.O. Box 1051
;        CITY:  Rochester
;        STATE:  New York
;        COUNTRY:  U.S.A.
;        ZIP:  14603
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE:  Floppy disk
;        COMPUTER:  IBM PC compatible
;        OPERATING SYSTEM:  PC-DOS/MS-DOS
;        SOFTWARE:  Patent In Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER:  US/08/891,254
;        FILING DATE:  10-JUL-1997
;        CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;        APPLICATION NUMBER:  08/475,775
;        FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;        NAME:  Goldman, Michael L.
;        REGISTRATION NUMBER:  30,727
;        REFERENCE/DOCKET NUMBER:  14603/10050
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE:  (716) 263-1304
;        TELEFAX:  (716) 263-1600
; INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH:  385 amino acids
;        TYPE:  amino acid
;        STRANDEDNESS:
;        TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-891-254-3

Alignment Scores:
Pred. No.:              9.24e-168       Length:         385
Score:                  1986.00         Matches:        385
Percent Similarity:     99.74%          Conservative:   0
Best Local Similarity:  99.74%          Mismatches:     0
Query Match:            78.44%          Indels:         1
DB:                     1               Gaps:           0

SEQ12-SEQ4 (1-1390) x US-08-891-254-3 (1-385)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCCGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCTAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGGCTGGGCGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTTGCCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy     355 ACCGGCATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy     415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluLeuSerAsn 100

Qy     475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGCCGGCAACAAT 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy     535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy     595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy     655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy     715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy     775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATCGTCTGAGCCAGCTCCTTGGCAAC 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuGlyAsn 220

Qy     835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy     895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy     955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGCGCTGAATGATATCGGTACGCAC 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy    1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGCGATCGGGCGATGGCGAAGGAAATC 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy    1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy    1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGCCAAAGCACTGAGCAAGCCAGATGAC 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy    1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy    1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCAGCGGGTGCCGGTGGTTCTTCG 1314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGln-HisAlaValProValValLeuAr 380

Qy    1315 CTGGGTAATGATGCCA 1330
           ||||||||||||||||
Db     380 gTrpValLeuMetPro 385

RESULT 7
US-08-819-539-3
; Sequence 3, Application US/08819539

; Patent No. 5859324
;  GENERAL INFORMATION:
;    APPLICANT:  Wei, Zhong-Min
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  Hypersensitive Response
;    TITLE OF INVENTION:  Induced Resistance In Plants
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/819,539
;      FILING DATE:  17-MAR-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/475,775
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  14603/10050
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  385 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-819-539-3

Alignment Scores:
Pred. No.::              9.24e-168       Length:          385
Score:                   1986.00         Matches:         385
Percent Similarity:      99.74%          Conservative:    0
Best Local Similarity:   99.74%          Mismatches:      0
Query Match:             78.44%          Indels:          1
DB:                      2               Gaps:            0

SEQ12-SEQ4 (1-1390) x US-08-819-539-3 (1-385)

Qy     175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy     235 GGCGGAAATAACGGGTTGCTCGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 GlyGlyAsnAsnGlyLeuGlyLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyAsnSer 40

Qy     295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy     355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy     415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluLeuSerAsn 100

Qy     475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGCCGGCAACAAT 534
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy      535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy      595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy      655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy      715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy      775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy      835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy      895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy      955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy     1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy     1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy     1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAGCACTGAGCAAGCCAGATGAC 1194
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy     1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy     1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
            ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGln-HisAlaValProValValLeuAr 380

Qy     1315 CTGGGTATTGATGCCA 1330
            ||||||||||||||||
Db      380 gTrpValLeuMetPro 385
```

RESULT 8
PCT-US96-08819-3
; Sequence 3, Application PC/TUS9608819
; GENERAL INFORMATION:
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
; TITLE OF INVENTION: RESISTANCE IN PLANTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/08819
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/475,775
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/10051
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 385 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-08819-3

Alignment Scores:

Pred. No.:	9.24e-168	Length:	385
Score:	1986.00	Matches:	385
Percent Similarity:	99.74%	Conservative:	0
Best Local Similarity:	99.74%	Mismatches:	0
Query Match:	78.44%	Indels:	1
DB:	5	Gaps:	0

SEQ12-SEQ4 (1-1390) x PCT-US96-08819-3 (1-385)

```
Qy      175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy      235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy      295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy      355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy      415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy      475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy      535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy      595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy      655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy    1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy    1075 GGTCAGTCAATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy    1135 GCTCAGGAGGTGAAAACCGATGACAAATCATGGCCAAAAGCACTGAGCAAGCCAGATGAC 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy    1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy    1255 CCCATGCGGGGTGATACCGGCAACGCACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGln-HisAlaValProValValLeuArg 380

Qy    1315 CTGGGTATTGATGCCA 1330
           ||||||||||||||||
Db    380 GTrpValLeuMetPro 385

RESULT 9
PCT-US93-06243-2
; Sequence 2, Application PC/TUS9306243
;  GENERAL INFORMATION:
;    APPLICANT:  Zhong-Min Wei, David W. Bauer, Steven V.
;    APPLICANT:  Beer, Alan Collmer, Sheng-Yang He, and Ron J. Laby
;    TITLE OF INVENTION:  Elicitor of the Hypersensitive Response in Plants
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Yahwak & Associates
;      STREET:  25  Skytop Drive
;      CITY:  Trumbull
;      STATE:  Connecticutt
;      COUNTRY:  USA
;      ZIP:  06611
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/06243
;      FILING DATE:  19930630
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  907,935
;      FILING DATE:  01-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  George M. Yahwak
;      REGISTRATION NUMBER:  26,824
;      REFERENCE/DOCKET NUMBER:  CRF D-1172
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (203)268-1951
;      TELEFAX:  (203)268-1951
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  385 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US93-06243-2

Alignment Scores:
Pred. No.:              1.23e-165       Length:         385
Score:                  1962.00         Matches:        381
Percent Similarity:     98.96%          Conservative:   1
Best Local Similarity:  98.70%          Mismatches:     3
Query Match:            77.49%          Indels:         1
DB:                     5               Gaps:           0

sEQ12-SEQ4 (1-1390) x PCT-US93-06243-2 (1-385)

Qy    175 ArGAGTCTGAATACAAGTGGGCTGGGACCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTAACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGGCTGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy    355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTGCAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuGlnAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240
```

```
Qy     895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
           |||||||||||||||:::   ||||||||||||||||||||||||||||||||||||||||
Db     241 GlyGlyLysGlyLeuArgGlyLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy     955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy    1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy    1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy    1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy    1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy    1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
           ||||||||||||||||||||||||||||||||||      ||||||||||||||||||||
Db     361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeu-HisAspAlaValProValValLeuAr 380

Qy    1315 CTGGGTATTGATGCCA 1330
           ||||||||||||||||
Db     380 gTrpValLeuMetPro 385
```

RESULT 10
US-08-891-254-1

```
; Sequence 1, Application US/08891254
; Patent No. 5776889
;  GENERAL INFORMATION:
;    APPLICANT:  Wei, Zhong-Min
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  Hypersensitive Response
;    TITLE OF INVENTION:  Induced Resistance In Plants
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/891,254
;      FILING DATE:  10-JUL-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/475,775
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  14603/10050
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-891-254-1

Alignment Scores:
Pred. No.:              1.25e-55      Length:        338
Score:                  718.50        Matches:       173
Percent Similarity:     52.97%        Conservative:  41
Best Local Similarity:  42.82%        Mismatches:    111
Query Match:            28.38%        Indels:        79
DB:                     1             Gaps:          11

SEQ12-SEQ4 (1-1390) x US-08-891-254-1 (1-338)

Qy     211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
           |||||||||:::|||            ||||||:::   |||:::   |||
Db       1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy     268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGCTGGGCGGCGGT 315
           |||      ||||||   |||   |||||||||          ||||||
Db      21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy     316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
              :::    |||::::::::|||      ||||||    ||||||
Db      39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy     376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                 ||||||   |||       |||
Db      56 -----------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy     436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
           ||||||   |||          |||   |||:::||||||   |||          |||
Db      64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy     496 GGTTCGCTGAACACGCTGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                             |||:::::|||   :::|||    |||
Db      83 Asn--------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy     556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                  ||||||
Db      97 ------------------------------------------------------SerGly 98

Qy     616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
              |||:::   |||                              |||||||||   :::
Db      99 GlyAspAlaLeuSer------------------------------LysMetPheAspLys 108

Qy     676 ATAATGCAAAGCCTGTTTGGT-----------------------GATGGCAAGATGGC 711
              :::       |||   |||                         :::       :::
Db     109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy     712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                    |||                 |||   |||:::|||       ||||||:::
Db     129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy     772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
              ||||||   :::|||||||||   ::::::|||||||||||||   |||   :::
Db     149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy     832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                 :::|||          |||
Db     168 -------------------------------------SerGlyPheSerGlnProSer 174

Qy     892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
           ||||||      |||||||||   |||||||||          :::   ||||||||||||
Db     175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy     952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
           |||:::|||   |||:::|||         |||   :::   ||||||:::::::::   |||
Db     195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy    1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
           |||          :::   |||   ||||||:::|||   ||||||   |||||||||||
```

```
Db     215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy    1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
           ||||||||||||||||||||||||||||||:::||||||||||||:::||||||||
Db     235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy    1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                                |||||||||||||||||||||||||||||||||||||||||
Db     255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy    1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
           |||||||||||||||  |||||||||::::::|||    :::|||   ||||||||||||
Db     275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy    1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
                 :::||||||||||||||||||   ||||||     ||||||||||||||:::
Db     295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy    1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
           |||||||||||||||||   :::   ||||||   |||   ||||||:::||||||||||
Db     315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy    1372 CTGGGCGCGGCT 1383
           |||   |||
Db     335 LeuAlaAsnAla 338
```

```
RESULT 11
US-08-484-358-2
; Sequence 2, Application US/08484358
; Patent No. 5850015
;  GENERAL INFORMATION:
;    APPLICANT:  Bauer, David
;    APPLICANT:  Collmer, Alan
;    TITLE OF INVENTION:  Hypersensitive Response Elicitor
;    TITLE OF INVENTION:  From
;    TITLE OF INVENTION:  Erwinia Chrysanthemi
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,358
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/840
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1304
;      TELEFAX:  716-263-1600
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-484-358-2

Alignment Scores:
Pred. No.:             1.25e-55      Length:          338
Score:                 718.50        Matches:         173
Percent Similarity:    52.97%        Conservative:    41
Best Local Similarity: 42.82%        Mismatches:      111
Query Match:           28.38%        Indels:          79
DB:                    2             Gaps:            11

SEQ12-SEQ4 (1-1390) x US-08-484-358-2 (1-338)
```

```
Qy     211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
           ||||||||:::|||      |||||||:::   |||:::    |||
Db       1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy     268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
           |||      ||||||   |||   ||||||||||         ||||||
Db      21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy     316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                :::   |||::::::::|||             ||||||   ||||||
Db      39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy     376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                  ||||||   |||     |||
Db      56 ---------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy     436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
           ||||||   |||         |||   |||:::||||||   |||           |||
Db      64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy     496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                              |||:::::::|||   :::|||         |||
Db      83 Asn--------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy     556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                 ||||||
Db      97 ----------------------------------------------------SerGly 98

Qy     616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
              |||:::   |||                                 |||||||||   :::
Db      99 GlyAspAlaLeuSer-----------------------------LysMetPheAspLys 108

Qy     676 ATAATGCAAAGCCTGTTTGGT----------------------GATGGGCAAGATGGC 711
              :::      |||   |||                          :::      :::
Db     109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy     712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                               |||         |||   |||:::|||      ||||||:::
Db     129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy     772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
              ||||||   :::|||||||||   ::::::||||||||||||   |||   :::
Db     149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy     832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                         :::|||         |||
Db     168 ---------------------------------------SerGlyPheSerGlnProSer 174

Qy     892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
           ||||||      |||||||||   |||||||||            :::   ||||||||||||
Db     175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy     952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
           |||:::|||   |||:::|||         |||   :::   ||||||:::::::::   |||
Db     195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy    1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
           |||         :::   |||   ||||||:::|||   ||||||   ||||||||||||
Db     215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy    1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
           ||||||||||||||||||||||||||||||:::||||||||||||:::||||||||
Db     235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy    1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                                |||||||||||||||||||||||||||||||||||||||||
Db     255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274
```

```
Qy   1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
          ||||||||||||||  ||||||||::::::|||   :::|||  |||||||||||
Db    275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy   1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
             :::||||||||||||||||||   ||||||      |||||||||||||||:::
Db    295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy   1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
          ||||||||||||||||||   :::   ||||||   |||   ||||||:::||||||||
Db    315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy   1372 CTGGGCGCGGCT 1383
          |||      |||
Db    335 LeuAlaAsnAla 338
```

RESULT 12
US-08-819-539-1

```
; Sequence 1, Application US/08819539
; Patent No. 5859324
;  GENERAL INFORMATION:
;    APPLICANT:  Wei, Zhong-Min
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  Hypersensitive Response
;    TITLE OF INVENTION:  Induced Resistance In Plants
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/819,539
;      FILING DATE:  17-MAR-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/475,775
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  14603/10050
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-819-539-1
```

Alignment Scores:

Pred. No.:	1.25e-55	Length:	338
Score:	718.50	Matches:	173
Percent Similarity:	52.97%	Conservative:	41
Best Local Similarity:	42.82%	Mismatches:	111
Query Match:	28.38%	Indels:	79
DB:	2	Gaps:	11

SEQ12-SEQ4 (1-1390) x US-08-819-539-1 (1-338)

```
Qy    211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
          |||||||||:::|||         ||||||:::   |||:::   |||
Db      1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy    268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
          |||     ||||||   |||   |||||||||      ||||||
Db     21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy    316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
             :::   |||:::::::::|||            ||||||   ||||||
Db     39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy    376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                 ||||||   |||      |||
Db     56 ----------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy    436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
          ||||||   |||         |||   |||:::||||||   |||            |||
Db     64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy    496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                            |||::::::|||   :::|||      |||
Db     83 Asn--------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy    556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                               ||||||
Db     97 ---------------------------------------------------SerGly 98

Qy    616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
             |||:::   |||                             |||||||||   :::
Db     99 GlyAspAlaLeuSer-----------------------------LysMetPheAspLys 108

Qy    676 ATAATGCAAAGCCTGTTTGGT-----------------------GATGGGCAAGATGGC 711
             :::      |||   |||                        :::      :::
Db    109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy    712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                      |||               |||   |||:::|||      ||||||:::
Db    129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy    772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
             ||||||   :::|||||||||   ::::::|||||||||||   |||   :::
Db    149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy    832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                 :::|||            |||
Db    168 ---------------------------------------SerGlyPheSerGlnProSer 174

Qy    892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
          ||||||      |||||||||   |||||||||         :::   ||||||||||||
Db    175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy    952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
          |||:::|||   |||:::|||      |||   :::   ||||||:::::::::   |||
Db    195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy   1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
          |||      :::   |||   |||||||:::|||   ||||||   ||||||||||||
Db    215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy   1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
          ||||||||||||||||||||||||||||||:::|||||||||||||:::|||||||||
Db    235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy   1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                   ||||||||||||||||||||||||||||||||||||||||||||||||
Db    255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy   1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
          ||||||||||||||  ||||||||::::::|||   :::|||  |||||||||||
Db    275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy   1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
             :::||||||||||||||||||   ||||||      |||||||||||||||:::
```

```
Db     295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy    1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
           |||||||||||||||||    :::    ||||||    |||    ||||||:::||||||||
Db     315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy    1372 CTGGGCGCGGCT 1383
           |||      |||
Db     335 LeuAlaAsnAla 338

RESULT 13
US-09-030-270A-1
; Sequence 1, Application US/09030270A
; Patent No. 5977060
;  GENERAL INFORMATION:
;    APPLICANT:  Zitter, Thomas A.
;    APPLICANT:  Wei, Zhong-Min
;    TITLE OF INVENTION:  INSECT CONTROL WITH A
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,270A
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/039,226
;      FILING DATE:  28-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1521
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-030-270A-1

Alignment Scores:
Pred. No.:              1.25e-55      Length:         338
Score:                  718.50        Matches:        173
Percent Similarity:     52.97%        Conservative:   41
Best Local Similarity:  42.82%        Mismatches:     111
Query Match:            28.38%        Indels:         79
DB:                     2             Gaps:           11

SEQ12-SEQ4 (1-1390) x US-09-030-270A-1 (1-338)

Qy     211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
           ||||||||:::|||                   ||||||:::    |||:::    |||
Db       1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy     268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
           |||      ||||||   |||   |||||||||                 ||||||
Db      21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy     316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
              :::    |||:::::::::|||                ||||||    ||||||
Db      39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy     376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                  ||||||    |||      |||
Db      56 ----------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy     436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
           ||||||   |||             |||    |||:::||||||   |||          |||
Db      64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy     496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                                   |||::::::|||    :::|||       |||
Db      83 Asn---------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy     556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                 ||||||
Db      97 -----------------------------------------------------SerGly 98

Qy     616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
              |||:::    |||                                 |||||||||    :::
Db      99 GlyAspAlaLeuSer--------------------------------LysMetPheAspLys 108

Qy     676 ATAATGCAAAGCCTGTTTGGT-----------------------GATGGGCAAGATGGC 711
              :::       |||    |||                          :::       :::
Db     109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy     712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                        |||                |||    |||:::|||       ||||||:::
Db     129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy     772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
              ||||||    :::|||||||||    ::::::||||||||||||    |||     :::
Db     149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy     832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                      :::|||          |||
Db     168 --------------------------------------SerGlyPheSerGlnProSer 174

Qy     892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
           ||||||       |||||||||    |||||||||           :::    |||||||||||
Db     175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy     952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
           |||:::|||    |||:::|||       |||    :::    ||||||:::::::::    |||
Db     195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy    1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
           |||          :::    |||    ||||||:::|||    ||||||    |||||||||||
Db     215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy    1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
           ||||||||||||||||||||||||||||||:::||||||||||||:::|||||||||
Db     235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy    1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy    1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
           ||||||||||||||    |||||||||::::::|||    :::|||    |||||||||||
Db     275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy    1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
               :::||||||||||||||||||    ||||||       |||||||||||||||:::
Db     295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy    1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
           |||||||||||||||||    :::    ||||||    |||    ||||||:::||||||||
Db     315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy    1372 CTGGGCGCGGCT 1383
```

```
               |||       |||
Db     335 LeuAlaAsnAla 338

RESULT 14
US-09-118-959-2
; Sequence 2, Application US/09118959
; Patent No. 6001959
;  GENERAL INFORMATION:
;    APPLICANT:  Bauer, David
;    APPLICANT:  Collmer, Alan
;    TITLE OF INVENTION:  Hypersensitive Response Elicitor From
;    TITLE OF INVENTION:  Erwinia Chrysanthemi
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/118,959
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/840
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-263-1304
;      TELEFAX:  716-263-1600
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-118-959-2

Alignment Scores:
Pred. No.:              1.25e-55     Length:         338
Score:                  718.50       Matches:        173
Percent Similarity:     52.97%       Conservative:   41
Best Local Similarity:  42.82%       Mismatches:     111
Query Match:            28.38%       Indels:         79
DB:                     3            Gaps:           11

SEQ12-SEQ4 (1-1390) x US-09-118-959-2 (1-338)

Qy     211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
           ||||||||:::|||            ||||||:::    |||:::    |||
Db       1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy     268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
           |||      ||||||   |||   |||||||||          ||||||
Db      21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy     316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                 :::   |||:::::::|||           ||||||   ||||||
Db      39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy     376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                ||||||   |||      |||
Db      56 ------------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy     436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
           ||||||   |||          |||   |||:::||||||   |||            |||
Db      64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy     496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                                    |||:::::|||   :::|||      |||
Db      83 Asn--------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy     556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                 ||||||
Db      97 ---------------------------------------------------SerGly 98

Qy     616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
              |||:::   |||                                |||||||||   :::
Db      99 GlyAspAlaLeuSer-----------------------------LysMetPheAspLys 108

Qy     676 ATAATGCAAAGCCTGTTTGGT-----------------------GATGGGCAAGATGGC 711
              :::      |||   |||                           :::      :::
Db     109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy     712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                    |||                |||   |||:::|||      ||||||:::
Db     129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy     772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
              ||||||   :::|||||||||   ::::::||||||||||||   |||   :::
Db     149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy     832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                 :::|||            |||
Db     168 -------------------------------------SerGlyPheSerGlnProSer 174

Qy     892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
           ||||||      |||||||||   |||||||||      :::   ||||||||||||
Db     175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy     952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
           |||:::|||   |||:::|||      |||   :::   ||||||:::::::::   |||
Db     195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy    1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
           |||         :::   |||   ||||||:::|||   ||||||   ||||||||||||
Db     215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy    1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
           ||||||||||||||||||||||||||||||:::||||||||||||:::|||||||||
Db     235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy    1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                             |||||||||||||||||||||||||||||||||||||||||||
Db     255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy    1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
           |||||||||||||||   ||||||||::::::|||   :::|||   ||||||||||||
Db     275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy    1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
                 :::||||||||||||||||||   ||||||      |||||||||||||||:::
Db     295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy    1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
           ||||||||||||||||||   :::   ||||||   |||   ||||||:::|||||||||
Db     315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy    1372 CTGGGCGCGGCT 1383
              |||     |||
Db     335 LeuAlaAsnAla 338

RESULT 15
US-08-984-207-1
; Sequence 1, Application US/08984207
; Patent No. 6235974
;  GENERAL INFORMATION:
;    APPLICANT:  Qiu, Dewen
;    APPLICANT:  Wei, Zhong-Min
```

```
;    APPLICANT:  Beer, Steven V.
;    TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE INDUCED
;    TITLE OF INVENTION:  RESISTANCE IN PLANTS BY SEED TREATMENT
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/984,207
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/033,230
;      FILING DATE:  05-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1201
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-984-207-1

Alignment Scores:
Pred. No.:              1.25e-55      Length:         338
Score:                  718.50        Matches:        173
Percent Similarity:     52.97%        Conservative:   41
Best Local Similarity:  42.82%        Mismatches:     111
Query Match:            28.38%        Indels:         79
DB:                     4             Gaps:           11

SEQ12-SEQ4 (1-1390) x US-08-984-207-1 (1-338)

Qy    211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
            |||||||||:::|||              ||||||:::   |||:::   |||
Db      1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy    268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
            |||      ||||||   |||   |||||||||         ||||||
Db     21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy    316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
               :::   |||::::::::|||          ||||||  ||||||
Db     39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy    376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                ||||||   |||      |||
Db     56 -----------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy    436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
            ||||||   |||          |||   |||:::||||||   |||         |||
Db     64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy    496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                             |||:::::|||   :::|||      |||
Db     83 Asn--------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy    556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                ||||||
Db     97 ---------------------------------------------------SerGly 98

Qy    616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
               |||:::   |||                            ||||||||    :::
Db     99 GlyAspAlaLeuSer----------------------------LysMetPheAspLys 108

Qy    676 ATAATGCAAAGCCTGTTTGGT----------------------GATGGGCAAGATGGC 711
               :::      |||   |||                         :::      :::
Db    109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy    712 ACCCAGGGCAGTTCCTCTGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                        |||              |||   |||:::|||      ||||||:::
Db    129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy    772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
               ||||||   :::|||||||||   ::::::||||||||||||   |||     :::
Db    149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy    832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                   .    :::|||       |||
Db    168 -------------------------------------SerGlyPheSerGlnProSer 174

Qy    892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
            ||||||      |||||||||   |||||||||         :::   ||||||||||||
Db    175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy    952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
            |||:::|||   |||:::|||      |||    :::   ||||||:::::::::   |||
Db    195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy   1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
            |||      :::   |||   ||||||:::|||   ||||||   ||||||||||||
Db    215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy   1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
            ||||||||||||||||||||||||||||||:::||||||||||||:::||||||||||
Db    235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy   1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                           ||||||||||||||||||||||||||||||||||||||||||||
Db    255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy   1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
            ||||||||||||||   ||||||||||::::::|||   :::|||   ||||||||||||
Db    275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy   1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
               :::||||||||||||||||||      ||||||      |||||||||||||||:::
Db    295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy   1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
            ||||||||||||||||    :::   ||||||   |||   ||||||:::||||||||||
Db    315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy   1372 CTGGGCGCGGCT 1383
            |||      |||
Db    335 LeuAlaAsnAla 338

Search completed: January 14, 2003, 17:19:34
Job time : 22.5 secs
```

GenCore version 5.1.3

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 14, 2003, 17:04:47 ; Search time 53.5 Seconds
(without alignments)
10706.747 Million cell updates/sec

Title: SEQ12-SEQ4
Perfect score: 2532
Sequence: 1 atgggatttttctcttttc.....gctgggcgcggcttaagctt 1390

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/DSPTO_spool/KUBELIK693/runat_14012003_091819_7076/app_query.fasta_1.1543
-DB=SPTREMBL_21 -OFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=KUBELIK693_@CGN_1_1_79_@runat_14012003_091819_7076 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -YGAPOP=6 -YGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	1181.5	46.7	382	2	Q9FCY8	Q9fcy8 erwinia ste
2	1009.5	39.9	365	2	Q9KH45	Q9kh45 pantoea agg
3	840.5	33.2	356	2	Q93FY8	Q93fy8 pectobacter
4	736.5	29.1	344	2	Q9EXP0	Q9exp0 erwinia chr
5	712	28.1	142	2	Q9LAW6	Q9law6 erwinia amy
c 6	298	11.5	763	2	Q9XDH2	Q9xdh2 mycobacteri
c 7	281.5	10.9	687	10	Q948Y7	Q948y7 volvox cart
c 8	279.5	10.8	598	16	Q8VKN7	Q8vkn7 mycobacteri
9	264.5	10.4	1489	16	O53559	O53559 mycobacteri
c 10	263	10.1	409	10	Q98BW1	Q98bw1 volvox cart
11	260.5	10.3	3659	16	Q98LN6	Q98ln6 rhizobium l
12	259	10.2	1665	16	O53215	O53215 mycobacteri
13	258	10.2	628	16	Q8VJ19	Q8vj19 mycobacteri
14	257.5	10.2	2174	16	Q92UU8	Q92uu8 rhizobium m
15	255.5	10.1	1381	16	O53552	O53552 mycobacteri
16	255	10.1	501	16	O07224	O07224 mycobacteri
17	255	10.1	1538	16	O53395	O53395 mycobacteri
18	254.5	10.1	622	16	Q8VKN3	Q8vkn3 mycobacteri
19	254	10.0	1079	16	O53557	O53557 mycobacteri
c 20	253.5	9.8	454	10	Q41645	Q41645 volvox cart
21	253.5	10.0	588	16	O50396	O50396 mycobacteri
c 22	253.5	9.8	744	10	O65375	O65375 arabidopsis
23	253	10.0	1507	16	Q8VJ23	Q8vj23 mycobacteri
c 24	251	9.7	368	10	Q09085	Q09085 phaseolus v
25	250	9.9	635	16	Q8VK71	Q8vk71 mycobacteri
26	250	9.9	775	16	Q8VJ15	Q8vj15 mycobacteri
27	249.5	9.9	1217	16	Q8VIY9	Q8viy9 mycobacteri
c 28	249.5	9.6	1315	10	Q9SPM0	Q9spm0 zea mays (m
29	249.5	9.9	1384	16	Q8VIZ1	Q8viz1 mycobacteri
30	249.5	9.9	1715	16	Q8VIZ0	Q8viz0 mycobacteri
31	249	9.8	731	16	O50415	O50415 mycobacteri
32	248	9.8	714	16	O53556	O53556 mycobacteri
c 33	245	9.5	489	10	Q41707	Q41707 vigna ungui
34	245	9.7	923	16	O53890	O53890 mycobacteri
c 35	241.5	9.3	386	10	Q9FPQ5	Q9fpq5 chlamydomon
36	241.5	9.5	879	16	Q8VKD2	Q8vkd2 mycobacteri
c 37	241	9.3	1188	10	Q41805	Q41805 zea mays (m
38	240.5	9.5	615	16	O05806	O05806 mycobacteri
39	240	9.5	1408	16	Q8VK17	Q8vk17 mycobacteri
c 40	239	9.2	760	10	Q9T0K5	Q9t0k5 arabidopsis
41	239	9.4	882	16	O53845	O53845 mycobacteri
c 42	237.5	9.2	280	10	Q43682	Q43682 vigna ungui
c 43	237.5	9.2	1611	3	O42854	O42854 schizosacch
44	236.5	9.3	436	16	Q8VIX6	Q8vix6 mycobacteri
45	236.5	9.3	853	16	O53439	O53439 mycobacteri

ALIGNMENTS

RESULT 1
Q9FCY8
ID Q9FCY8 PRELIMINARY; PRT; 382 AA.
AC Q9FCY8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HrpN.
GN HRPN.
OS Erwinia stewartii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=66271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS104;
RA Frederick R.D., Ahmad M., Majerczak D.R., Arroyo-Rodriguez A.S.,
RA Manulis S., Coplin D.L.;
RT "Genetic organization of the Pantoea stewartii subsp. stewartii hrp
RT gene cluster and sequence analysis of the hrpA, hrpC, hrpN and wtsE
RT operons.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF282857; AAG01466.1; -.
SQ SEQUENCE 382 AA; 37878 MW; 8BABC723F2DA388B CRC64;

```
Alignment Scores:
Pred. No.:               5.33e-75      Length:         382
Score:                   1181.50       Matches:        251
Percent Similarity:      73.02%        Conservative:   44
Best Local Similarity:   62.13%        Mismatches:     86
Query Match:             46.66%        Indels:         23
DB:                      2             Gaps:           11

SEQ12-SEQ4 (1-1390) x Q9FCY8 (1-382)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||:::||||||||  |||||   |||    :::|||:::::::::
Db      1 MetSerMetAsnThrSerProLeuGlyThrSerAlaLeuGlnValThrLeu--------- 17

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||:::||||||        :::     |||||||||      |||
Db     18 GlyGlyAsnAsnGlyLeuMetGlyThrAspLeuArgThrAspGlyLeuGlyLeuLeuSer 37

Qy    295 GCACTGGGGCTGGGC---GGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTA 351
                |||||||||    |||          |||::::::::::::    ||||||
Db     38 GlnProGlyLeuGlyGluGlyLysGlyHisAsnGluSerIleAspLeuLeuAlaAlaAla 57

Qy    352 CTCACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGC 411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     58 LeuThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeu---SerSerLeu 76

Qy    412 TTAGGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCG 471
          ||||||   |||   |||   |||                ||||||    |||
Db     77 LeuGlySerGlyThrGlyMetGlyAsnSerProPheGlyGlySerGlySerAlaProGly 96

Qy    472 AACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAAC 531
          |||   |||:::          |||                    ||||||   ||||||
Db     97 AsnThrLeuSerGlyThrSerGly------------------GlySerProGlyGly--- 109

Qy    532 AATACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCC 591
              ||||||         :::|||                ||||||::::::    |||
Db    110 ---ThrThrGlyAlaGlySerSer---------------LeuGlyLeuAspProThrGln 123

Qy    592 CAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAG 651
                |||||||||   ||||||         ||||||        |||   |||||||:::
Db    124 ThrGlyAspAspSerLeuSerGlyAlaGlyGlnThrSerGlyMetSer---ProMetGlu 142

Qy    652 CAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGC 711
          ||||||:::|||:::|||::::::|||   |||||||||||||||||||    |||||||||
Db    143 GlnLeuMetLysIlePheAlaAspIleThrGlnSerLeuPheGlyAsp---GlnAspGly 161

Qy    712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                |||     ::::::     |||::::||||||||:::::     ||||||||||||||||||
Db    162 AlaSerGlyGlyAsnAla---GlyArgGlnProSerGlnAspGluGlnAsnAlaTyrLys 180

Qy    772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
          |||||||||||||||||||||:::        ||||||   ||||||||||||:::      |||
Db    181 LysGlyValThrAspAlaLeuThrAlaPheMetGlyGlyGlyLeuSerGlnValAlaGly 200

Qy    832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
          ||||||      ||||||       ||||||       |||    |||       |||:::
Db    201 AsnGlySerGluGlyGlyLeuAspGlyGly---MetGlyLeuGly---GlyGlyAsnGly 218

Qy    892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
          ||||||||||||||||||||||:::||||||||||||||||||||    |||:::||||||||||||||||
Db    219 LeuGlyGlyLysGlyLeuGlnAspLeuSerGlyProAlaAspPheGlnGlnLeuGlyAsn 238

Qy    952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
          |||:::|||||||||:::|||||||||||||||||||:::||||||||||:::|||||||||
Db    239 AlaIleGlyThrGlyValGlyMetLysAlaGlyIleGluAlaLeuAsnAsnIleGlyThr 258

Qy   1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
          |||       ||||||||||||||||||:::||||||     |||||||||:::|||:::|||
Db    259 HisSerAspSerSerThrArgSerPheIleAsnLysGluAspArgAlaLeuAlaArgGlu 278

Qy   1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
          :::|||||||||||||||||||||||||||     ||||||||||||||||||||||||
Db    279 ValGlyGlnPheMetAspGlnTyrProGluThrPheGlyLysProGlnTyrGlnLysAsn 298

Qy   1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                      ||||||||||||     ||||||||||||:::|||||||||:::||||||
Db    299 AlaAspSerAlaValLysThrAspThrLysSerTrpAlaGluAlaLeuSerGlnProAsp 318

Qy   1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
          ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db    319 AspAspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyIleIleLys 338

Qy   1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
                ||||||||||||   ||||||     ||||||||||||||||||||||||||||||
Db    339 SerAlaMetAlaGlyAspAsnGlyAsnIleAsnLeuGlnAlaArgGlyAlaGlyGlySer 358

Qy   1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
          |||:::|||||||||||||     :::     |||||||||||||||||||||||||     :::
Db    359 SerMetGlyIleAspAlaThrLeuThrGlyAspAlaIleAsnAsnMetAlaLeuThrArg 378

Qy   1372 CTGGGCGCGGCT 1383
          |||   ||||||
Db    379 LeuSerAlaAla 382

RESULT 2
Q9KH45
ID   Q9KH45      PRELIMINARY;      PRT;   365 AA.
AC   Q9KH45;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Harpin.
GN   HRPN.
OS   Pantoea agglomerans pv. gypsophilae.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pantoea.
OX   NCBI_TaxID=48984;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21171042; PubMed=11277443;
RA   Mor H., Manulis S., Zuck M., Nizan R., Coplin D.L., Barash I.;
RT   "Genetic organization of the hrp gene cluster and dspAE/BF operon in
RT   Erwinia herbicola pv. gypsophilae.";
RL   Mol. Plant Microbe Interact. 14:431-436(2001).
DR   EMBL; AF271716; AAF76342.1; -.
SQ   SEQUENCE   365 AA;  36784 MW;  6F79BB6DDD95AC2A CRC64;

Alignment Scores:
Pred. No.:               6.5e-63       Length:         365
Score:                   1009.50       Matches:        224
Percent Similarity:      67.51%        Conservative:   42
Best Local Similarity:   56.85%        Mismatches:     90
Query Match:             39.87%        Indels:         38
DB:                      2             Gaps:           13

SEQ12-SEQ4 (1-1390) x Q9KH45 (1-365)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          |||||||||||||||:::   ||||||          ||||||||||||:::
Db      1 MetSerLeuAsnThrThrProLeuGlyIleProAlaMetGlnIleSerLeu--------- 17

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          |||   :::|||||||||   |||   :::   |||||||||   |||
Db     18 GlyAspSerAsnGlyLeuProGlyProAsnLeuGlnAsnAlaValLeuAsn--------- 34

Qy    295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
                :::|||      :::   ::::::|||:::||||||||||||||   |||
Db     35 ------SerMetGlyAspArgSerThrAspGluThrIleAsnGlnLeuAlaGlyAlaLeu 52

Qy    355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          |||   |||||||||||||||:::||||||||||||   ||||||   ||||||||||||
Db     53 ThrAlaMetMetMetMetMetAsnMetMetGlyGlySerGlyLeu---GlyGlyGlyLeu 71

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGC------TCAGGTGGCCTGGGCGAAGGACTG 468
          |||         |||   |||||||||         |||   ||||||      |||
Db     72 GlyThrAla---GlyGlyGluLeuGlyGlyGlnSerGlyGlyAlaLeuGlyGlyProLeu 90
```

```
Qy    469 TCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACG---CTGGGCTCGAAAGGC 525
                      |||||||||:::|||    |||    ||||||        |||
Db     91 TyrGlyAsnGlnGlyGlyGlnLeuGlyGlyAlaLeuGlyThrSerLeuGlyGlyThrGly 110

Qy    526 GGCAACAATACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCA 585
           |||                        |||    ||||||:::||||||    :::
Db    111 Gly-----------------AlaGlySerSerLeuAspLysAlaLeuAspLeu------ 122

Qy    586 ACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGC------ 639
                         |||   |||:::             |||||||||:::
Db    123 ---------------AspProThrAlaAspGlyGlyGlyLeuSerAspSerAsnGlyThr 137

Qy    640 ---------GACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTG 690
                         |||:::   ||||||:::|||||||||:::|||:::|||||||||:::
Db    138 TyrGlyMetSerProIleAspGlnLeuMetLysMetPheAlaGluValMetGlnSerMet 157

Qy    691 TTTGGTGATGGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAA 750
           ||||||   ||||||       |||   |||       :::   |||   |||||||||
Db    158 PheGly---GlyGlnGlyAspThrSerGlyArgGlyThr---GlyAsnGlnProThrPro 175

Qy    751 GGCGAGCAGAACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAAT 810
                |||||||||||||||   |||||||||||||||||||||:::      :::||||||
Db    176 AspGluGlnAsnAlaTyrThrLysGlyValThrAspAlaLeuThrAlaValMetGlyGly 195

Qy    811 GGTCTGAGCCAGCTCCTTGGCAACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGC 870
           ||||||||||||:::   |||:::   |||   |||||||||        |||:::   |||
Db    196 GlyLeuSerGlnMetGlnGlySer---GlyThrGlyGlyGlyMetAsnGlySerIleGly 214

Qy    871 ACGGGTCTTGACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTG 930
                         |||:::   ||||||||||||||||||||||   |||:::||||||
Db    215 Leu---------GlyAsnGlyLeuGlyGlyLysGlyLeuGlnGluLeuAsnGlyProAla 231

Qy    931 GACTACCAGCAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAG 990
           ||||||:::|||||||||         ||||||||||||:::||||||||||||||||||:::
Db    232 AspTyrGluGlnLeuGlyValGlyValGlyThrGlyValGlyMetLysAlaGlyIleGlu 251

Qy    991 GCGCTGAATGATATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGC 1050
           |||||||||:::|||||||||         ||||||||||||:::|||:::||||||
Db    252 AlaLeuAsnAsnIleGlyThrGluSerAspSerSerThrArgThrPheIleAsnLysAsp 271

Qy   1051 GATCGGGCGATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGC 1110
           ||||||::::::|||||||||:::||||||||||||||||||||||||||||:::||||||
Db    272 AspArgSerLeuAlaLysGluValGlyGlnPheMetAspGlnTyrProGluIlePheGly 291

Qy   1111 AAGCCGCAGTACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCA 1170
           ||||||||||||||||||   |||         :::||||||||||||   ||||||||||||
Db    292 LysProGlnTyrGlnLysThrProTyrSerAspValLysThrAspThrLysSerTrpAla 311

Qy   1171 AAAGCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAAC 1230
           :::|||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db    312 GluAlaLeuSerAsnProAspAspAspGlyMetThrProAlaSerMetGluGlnPheAsn 331

Qy   1231 AAAGCCAAGGGCATGATCAAAAGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAG 1290
           |||||||||||||||||||||         |||   |||||||||||||||
Db    332 LysAlaLysGlyMetIleLysSerAlaMetGluGlyAspThrGlyAsnLeuGlnAlaThr 351

Qy   1291 GCACGCGGTGCCGGTGGTTCTTCGCTGGGTATTGATGCCA 1330
                |||   |||   :::   |||:::|||||||||
Db    352 GlyGluGlyAspGlyAsp-ValHisTrpLeuLeuMetPro 364
```

RESULT 3
Q93FY8

```
ID   Q93FY8     PRELIMINARY;     PRT;   356 AA.
AC   Q93FY8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HrpN.
GN   HRPN.
OS   Pectobacterium carotovorum subsp. carotovorum.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pectobacterium.
OX   NCBI_TaxID=555;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SCC1;
RA   Rantakari A., Virtaharju O., Vaehaemiko S., Taira S., Palva T.E.,
RA   Saarilahti H.T., Romantschuk M.;
RT   "Type III secretion contributes to the pathogenesis of soft-rotting
RT   plant pathogen Erwinia carotovora: Partial characterization of the hrp
RT   gene cluster.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF302656; AAK97283.1; -.
SQ   SEQUENCE   356 AA;  35534 MW;  154C9D885810DCB8 CRC64;

Alignment Scores:
Pred. No.:              4.9e-51     Length:        356
Score:                  840.50      Matches:       198
Percent Similarity:     57.49%      Conservative:  36
Best Local Similarity:  48.65%      Mismatches:    114
Query Match:            33.20%      Indels:        59
DB:                     2           Gaps:          11

SEQ12-SEQ4 (1-1390) x Q93FY8 (1-356)

Qy    181 CTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGA 240
           ||||||:::   |||   |||||||||      :::||||||:::|||      |||||||||
Db      2 LeuAsnSerLeuGlyGlyGlyAlaSer---LeuGlnIleThrIle---LysAlaGlyGly 19

Qy    241 AATAACGGGTTGCTG---GGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTGCA 297
           |||   ||||||         |||   ||||||   |||                ||||||
Db     20 AsnGlyGlyLeuPheProSerGlnSerSerGlnAsnGlyGlySerProSerGlnSerAla 39

Qy    298 CTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACC 357
                      ||||||   :::::::::            ||||||:::   ::::::|||
Db     40 Phe---------GlyGlyGlnArgSerAsnIleAlaGluGlnLeuSerAspIleMetThr 56

Qy    358 GGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGC 417
              ||||||   |||   ||||||||||||||||||          |||   ||||||||||||
Db     57 ThrMetMetPheMetGlySerMetMetGlyGlyGly------MetSerGlyGlyLeuGly 74

Qy    418 ------------------GGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGC 459
                            ||||||||||||   ||||||   |||   ||||||||||||
Db     75 GlyLeuGlySerSerLeuGlyGlyLeuGlyGlyGlyLeuLeuGly---GlyGlyLeuGly 93

Qy    460 GAAGGACTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCG 519
              ||||||   ::::::|||         ||||||   :::|||         |||
Db     94 GlyGlyLeuGlySerSerLeuGlySerGlyLeuGlySerAlaLeuGly------GlyGly 111

Qy    520 AAAGGCGGCAACAATACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATT 579
              ||||||                                          |||||||||
Db    112 LeuGlyGly---------------------------------------AlaLeuGlyAla 118

Qy    580 AACTCAACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGC 639
                 :::   |||      :::         |||:::
Db    119 GlyMetAsnAlaMetAsnProSerAlaMetMetGlySer--------------------- 131

Qy    640 GACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGAT 699
                               :::         ::::::   |||   |||
Db    132 -----------------------LeuLeuPheSerAlaLeuGluAspLeuLeuGlyGly 143

Qy    700 GGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAG 759
           |||               ||||||         |||   |||||||||:::         |||
Db    144 GlyMetSerGlnGlnGlnGlyGlyLeuPheGlyAsnLysGlnProSerSerProGluIle 163

Qy    760 AACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGC 819
           :::||||||   :::||||||   ||||||||||||   ::::::||||||||||||||||||
Db    164 SerAlaTyrThrGlnGlyValAsnAspAlaLeuSerAlaIleLeuGlyAsnGlyLeuSer 183

Qy    820 CAGCTCCTTGGCAACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTT 879
           |||      |||
Db    184 GlnThrLysGlyGln---------------------------------------Thr 189

Qy    880 GACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAG 939
                     ||||||      |||||||||   |||||||||            :::
```

```
Db      190 SerProLeuGlnLeuGlyAsnAsnGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsn 209

Qy      940 CAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAAT 999
            |||||||||:::   :::|||    :::|||   |||||||||:::|||   ||||||
Db      210 GlnLeuGlySerThrLeuGlyMetSerValGlyGlnLysAlaGlyLeuGlnGluLeuAsn 229

Qy     1000 GATATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCG 1059
            :::|||   ||||||      |||   ||||||   |||||||:::|||   ||||||
Db      230 AsnIleSerThrHisAsnAspSerProThrArgTyrPheValAspLysGluAspArgGly 249

Qy     1060 ATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAG 1119
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db      250 MetAlaLysGluIleGlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlu 269

Qy     1120 TACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTG 1179
            |||||||||         |||      |||   |||||||||||||||||||||||||||
Db      270 TyrGlnLysAspAsnTrpGlnThrAlaLysGlnAspAspLysSerTrpAlaLysAlaLeu 289

Qy     1180 AGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAG 1239
            ||||||||||||||||||||||||||         ||||||::::::|||   ||||||
Db      290 SerLysProAspAspAspGlyMetThrLysGlySerMetAspLysPheMetLysAlaVal 309

Qy     1240 GGCATGATCAAAAGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGT 1299
            |||||||||         :::||||||||||||||||||   ||||||   |||||||||
Db      310 GlyMetIleAsnThrAlaValAlaGlyAspThrGlyAsnThrAsnLeuSerAlaArgGly 329

Qy     1300 GCCGGTGGTTCTTCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATG 1359
            ||||||:::||||||||||||||||||   |||   ||||||   |||   ||||||
Db      330 AsnGlyGlyAlaSerLeuGlyIleAspAlaAlaMetIleGlyAspArgIleValAsnMet 349

Qy     1360 GCACTTGGCAAGCTGGGCGCG 1380
               |||   ||||||   :::
Db      350 GlyLeuLysLysLeuSerSer 356
```

RESULT 4

Q9EXP0

```
ID   Q9EXP0      PRELIMINARY;      PRT;   344 AA.
AC   Q9EXP0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Harpin protein.
GN   HRPN.
OS   Erwinia chrysanthemi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pectobacterium.
OX   NCBI_TaxID=556;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3937;
RA   Vedel R., Berthuis I., Boccara M.;
RT   "HrpN of Erwinia chrysanthemi 3937 Role in Pathogenicity and
RT   regulation.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ302027; CAC20124.1; -.
SQ   SEQUENCE   344 AA;  34493 MW;  137E535659D1CEF0 CRC64;
```

Alignment Scores:

Pred. No.:	9.96e-44	Length:	344
Score:	736.50	Matches:	171
Percent Similarity:	53.09%	Conservative:	44
Best Local Similarity:	42.22%	Mismatches:	111
Query Match:	29.09%	Indels:	79
DB:	2	Gaps:	9

SEQ12-SEQ4 (1-1390) x Q9EXP0 (1-344)

```
Qy      211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
            |||||||||:::|||         ||||||:::   |||:::   |||
Db        1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyLeu 20

Qy      268 CAGAATGCTGGGTTGGGTGGC------------AATTCTGCACTGGGGCTGGGCGGCGGT 315
                  ||||||   |||                |||:::||||||
Db       21 GlyAlaGlnGlyLeuLysGlyLeuAsnSerAlaThrSerSerLeuGlySerSerLeuAsp 40

Qy      316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
               :::   |||:::::::::|||            ||||||   ||||||
Db       41 LysLeuSerGlyThrIleAspLysLeuThrSerAlaLeuThrSerMetMetPhe------ 58

Qy      376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                              ||||||   :::   |||:::|||   ||||||      |||
Db       59 ---------------GlyGlyAlaLeuSerGlnGlyMetGlySerGlyLeuAlaArgGly 73

Qy      436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
               |||                                          :::   ||||||
Db       74 AlaGly------------------------------------------AsnGlnLeuGly 79

Qy      496 GGTTCGCTGAACACGCTGGGCTCG------AAAGGCGGCAACAATACCACTTCAACAACA 549
               |||||||||||   ||||||      :::|||   |||      |||
Db       80 AlaSerLeuAsnThrPheGlySerGlyAlaGlnGlyAlaGlyAsnValLeuSerLysPro 99

Qy      550 AATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACC 609

Db      100 Gln--------------------------------------------------------- 100

Qy      610 TCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTC 669
            ||||||:::|||:::   |||                              |||||||||
Db      101 SerGlySerAspAlaLeuSer------------------------------LysMetPhe 110

Qy      670 AGCGAGATAATGCAAAGCCTGTTTGGT------------------------GATGGGCAA 705
             :::   :::      |||   |||                           :::
Db      111 AspLysAlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSer 130

Qy      706 GATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCC 765
            :::               |||                |||   |||:::         ||||||
Db      131 AsnGlnLeuAlaAsnSerLeuLeuAsnAlaSerGlnMetThrGlnAlaAsnMetAsnAla 150

Qy      766 TATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTC 825
            :::      |||:::   ||||||||||||   :::::::||||||||||||   |||
Db      151 PheGlySerGlyLeuAsnAspAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnAla 170

Qy      826 CTTGGCAACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGT 885
            :::|||                                          |||
Db      171 MetGly------------------------------------------GlyPheSerPro 176

Qy      886 TCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTA 945
               |||||||||      |||||||||   |||:::|||            :::   ||||||
Db      177 LeuSerLeuGlyAlaGlyGlyLeuGlnGlyLeuAsnGlyAlaGlyAlaPheSerGlnLeu 196

Qy      946 GGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATC 1005
            |||||||||:::|||   |||:::|||      |||   :::   ||||||:::::::::
Db      197 GlyAsnAlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuAsnAlaLeuSerAsnVal 216

Qy     1006 GGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCG 1065
            ||||||         :::   |||:::||||||:::|||   ||||||   :::|||
Db      217 SerThrHisValAspGlyAsnAsnArgAsnPheValAspLysGluAspArgGlyLeuAla 236

Qy     1066 AAGGAAATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAG 1125
            |||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db      237 LysGluIleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGlnTyrGln 256

Qy     1126 AAAGGCCCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAG 1185
            |||               |||||||||||||||||||||||||||||||||||||||||||
Db      257 LysAspGlyTrpSerSerAlaLysThrAspAspLysSerTrpAlaLysAlaLeuSerLys 276

Qy     1186 CCAGATGACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATG 1245
            |||||||||||||||||||||   ||||||||::::::|||   :::|||   ||||||
Db      277 ProAspAspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaLeuGlyMet 296

Qy     1246 ATCAAAAGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGT 1305
            ||||||         :::||||||||||||||||||   ||||||      ||||||||||||
Db      297 IleLysSerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGly 316

Qy     1306 GGTTCTTCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTT 1365
            |||:::||||||||||||||||||   :::   ||||||   |||   ||||||:::|||
Db      317 GlyAlaSerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeu 336
```

```
Qy    1366 GGCAAGCTGGGCGCG 1380
           |||||||||   |||
Db     337 GlyLysLeuAlaAla 341

RESULT 5
Q9LAW6
ID   Q9LAW6      PRELIMINARY;      PRT;   142 AA.
AC   Q9LAW6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   Harpin HrpN (Fragment).
GN   HRPN.
OS   Erwinia amylovora.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Erwinia.
OX   NCBI_TaxID=552;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EA246;
RA   Kim J.F., Laby R.J., Beer S.V.;
RT   "Comparison of the hrpN-flanking regions of two Erwinia amylovora
RT   strains with different host specificity.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF083620; AAF63398.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   142 AA;  14690 MW;  7EF077EE743E589D CRC64;

Alignment Scores:
Pred. No.:              4.55e-42       Length:         142
Score:                  712.00         Matches:        138
Percent Similarity:     97.18%         Conservative:   0
Best Local Similarity:  97.18%         Mismatches:     4
Query Match:            28.12%         Indels:         0
DB:                     2              Gaps:           0

SEQ12-SEQ4 (1-1390) x Q9LAW6 (1-142)

Qy     958 GGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCACAGG 1017
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHisSer 20

Qy    1018 CACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATCGGT 1077
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 AspSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIleGly 40

Qy    1078 CAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGT 1137
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 GlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyProGly 60

Qy    1138 CAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGAC 1197
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAspAsp 80

Qy    1198 GGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGGCCC 1257
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysSerAla 100

Qy    1258 ATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCGCTG 1317
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 MetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSerLeu 120

Qy    1318 GGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTGGGC 1377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeuGly 140

Qy    1378 GCGGCT 1383
           ||||||
Db     141 AlaAla 142

RESULT 6
Q9XDH2
ID   Q9XDH2      PRELIMINARY;      PRT;   763 AA.
AC   Q9XDH2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Proline-rich mucin homolog.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=20090472; PubMed=10627046;
RA   Espitia C., Laclette J.P., Mondragon-Palomino M., Amador A.,
RA   Campuzano J., Martens A., Singh M., Cicero R., Zhang Y., Moreno C.;
RT   "The PE-PGRS glycine-rich proteins of Mycobacterium tuberculosis: a
RT   new family of fibronectin-binding proteins?";
RL   Microbiology 145:3487-3495(1999).
DR   EMBL; AF071081; AAD41594.1; -.
DR   InterPro; IPR002951; Atrophin.
DR   InterPro; IPR003882; Pistil_extensin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01222; ATROPHIN.
DR   PRINTS; PR01582; KV33CHANNEL.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
SQ   SEQUENCE   763 AA;  75035 MW;  39168EC45A5916F8 CRC64;

Alignment Scores:
Pred. No.:              7.52e-13       Length:         763
Score:                  298.00         Matches:        121
Percent Similarity:     37.08%         Conservative:   44
Best Local Similarity:  27.19%         Mismatches:     144
Query Match:            11.50%         Indels:         136
DB:                     2              Gaps:           24

SEQ12-SEQ4 (1-1390) x Q9XDH2 (1-763)

Qy    1379 GCGCCCAGCTTGCCAAGTGCCATATTGTTAATGGCATCACCGGCCATCATGGCATCAATA 1320
           |||||  :::     |||   :::|||:::|||              :::       :::
Db     268 AlaProProValAlaValAlaAlaValLeuValAlaProCysProProLeuProProLeu 287

Qy    1319 CCCAGCGAAGAACCACCGGCACCGCGTGCCTGCAGGTTGCCG------------------ 1278
           |||:::       |||||||||||  |||     :::|||
Db     288 ProAsnAsnHisProProAlaProProAlaAlaProValProGlyValProLeuAlaPro 307

Qy    1277 ------------------------TTGCCGGTATCACCCGCCATGGGCCTTTTGATCATG 1242
                                   |||   :::|||      |||:::    :::
Db     308 LeuProAsnSerHisProProAlaProProSerAlaProValProGlyValProLeuAla 327

Qy    1241 CCCTTGGCTTTGTTGAACTGCTCCATACTGGCTGGTGTCATTCCGTCGTCATCTGGCTTG 1182
           ||||||    :::                 :::           |||        |||
Db     328 ProLeuProIleSerGlyArgProVal------SerValTrpLysGlySerPheThrThr 345

Qy    1181 CTCAGTGCTTTTGCCCATGATTTGTCATCGGTTTTCACCTCCTGACCCGGGCCTTTCTGG 1122
           ||||||    |||          :::
Db     346 LeuSerThrPheCysCysArgVal------------------------------------ 353

Qy    1121 TACTGCGGCTTGCCAAACACCTCAGGATACTGGTCCATGAACTGACCGATTTCCTTCGCC 1062
                 |||           :::|||      :::    :::   |||
Db     354 ---CysSerGlyGluValLeuAlaGlyAlaLeuAsnProSerArgPro------------ 368

Qy    1061 ATCGCCCGATCGCCTTTATTGACGAAAGAACGGGTTGAACTGTGCCTGTGCGTACCGATA 1002
              :::||||||||||||   |||                      |||      ||||||
Db     369 ---SerArgSerProLeuThrThr---------ThrThrProAlaLeuProAlaProIle 384

Qy    1001 TCATTCAGCGCCTGAATGCCCGCTTTCATACCGATACCGGTACCCACGGCGTTACCTAAC 942
                             :::|||         |||:::|||:::   |||||||:::|||
Db     385 ---------ProProLeuProProLeuProProLeuProIleAsnThrAlaValPro--- 400

Qy     941 TGCTGGTAGTCCACCGGCCCGCTCAGGTTTTGCAGCCCTTTGCCGCCCAGCGACGAACCG 882
                                 |||:::              |||||||||||
```

```
Db    401 -----------------ProIle---------ProProLeuProProValThrAlaLeu 411

Qy    881 TCAAGACCCGTGCCAGCATTACCGCCCTGACCA--------------CCTCCCAGTCCC 837
          :::  |||:::|||  |||  |||  |||              |||||||:::|||
Db    412 AlaProProLeuProProLeuAlaProLeuProIleSerProGlyValProProAlaPro 431

Qy    836 CCGTTGCCA------AGGAGCTGG-----------CTCAGACCATTACCCATCAGGCCC 795
          |||:::|||      :::  |||            |||  |||  |||      |||
Db    432 ProIleProProGlyLysProTrpThrThrProProLeuAlaProAlaProProGluPro 451

Qy    794 GACAGCGCATCAGTGACTCCTTTTTATAGGCGTTCTGCTCGCCTTCGGTCGGCTGCTTG 735
             :::      |||   |||                    |||         |||
Db    452 LysThrValProValLeuPro--------------------ProGlyProSerCys--- 463

Qy    734 CCCCCAGAGGAACTGCCCTGGGTGCCATCTTGCCCATCACCAAACAGGCTTTGCATTATC 675
          ||||||  |||  |||    |||:::   |||  |||
Db    464 ProProSerGluLysProAsnProProAlaProProGluPro----------------- 477

Qy    674 TCGCTGAACATCTTCAGCAGCTGCTGCATCGGGTCGCTGGAGTCTGAGGTGGAATCTGTG 615

Db    477 ------------------------------------------------------------ 477

Qy    614 CCGGAGGTGGAATCGTCGTTTTGGGACGTTGAGTTAATACCCAGCGCCTGGTCCAGCGGG 555
          ||||||   :::||||||                 :::|||   |||
Db    478 ProGluProLysSerSer------------ProAlaLeuProProAla---ProProAla 492

Qy    554 GAATTTGTTGTTGAAGTGGTATTGTTGCCGCCTTTCGAGCCCAGCGTGTTCAGCGAACCG 495
             :::         |||   :::||||||      |||   :::              |||
Db    493 ProSerMetProSerAlaValArgValProPro---SerProProIleProProAlaPro 511

Qy    494 CCTAACATATCGTTCAGCGCGTTCGACAGTCCTTCGCCCAGGCCACCTGAGCCACCC--- 438
          |||            ::::::   :::      |||   ||||||   ||||||
Db    512 ProAlaAlaProArgAlaSerMetProAlaLeuProProAlaProProSerProProAla 531

Qy    437 ------------AAGCCATTACCTAAGCCACCGCCTAAGCCA---------------CCG 405
                      |||||||||   ||||||   |||                  |||
Db    532 ThrArgLeuCysProProLeuProProSerProProAlaProAsnSerProProAlaPro 551

Qy    404 CCCATCAGCCCACCACCGCCCATCATGCTCATC--------------------ATCATC 366
          |||    |||   ||||||   :::|||                         :::
Db    552 ProAlaProProThrProProLysLeuLeuSerAlaAsnProProCysProProValPro 571

Qy    365 ATCATGCCGGTGAGTAAGCCAGCCAGCTGATTGACGGTATCATTTTGATTACCGCCGCCC 306
                |||         ||||||                          ||||||
Db    572 ProAlaProAsnArgProProAla----------------------------ProPro 581

Qy    305 AGCCCCAGTGCA------GAATTG---------------CCACCCAACCCAGCATTCTGG 267
          :::|||    |||     ||||||                |||||| |||
Db    582 AlaProProAlaProProGluLeuProAlaProProAspProProThrProProValAla 601

Qy    266 CGACTGGTACCCAGCAACCCGTTATTTCCGCCCGCACCGCCGATAGAAATTTGCATCGTT 207
                  |||:::   |||      ||||||||||||||    :::       |||
Db    602 AsnSerProProAlaProProAla---ProProAlaProProSerAlaLeuProPheVal 620

Qy    206 GACGCTCCCAGCCCA 192
          :::   |||:::|||
Db    621 AsnProProAlaPro 625
```

RESULT 7
Q948Y7

```
ID   Q948Y7      PRELIMINARY;      PRT;    687 AA.
AC   Q948Y7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   VMP3 protein.
GN   VMP3.
OS   Volvox carteri f. nagariensis.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Volvocaceae; Volvox.
OX   NCBI_TaxID=3068;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HK 10;
RX   MEDLINE=21382906; PubMed=11489172;
RA   Hallmann A., Amon P., Godl K., Heitzer M., Sumper M.;
RT   "Transcriptional activation by the sexual pheromone and wounding: a
RT   new gene family from Volvox encoding modular proteins with
RT   (hydroxy)proline-rich and metalloproteinase homology domains.";
RL   Plant J. 26:583-593(2001).
DR   EMBL; AJ311549; CAC39318.1; -.
DR   MEROPS; M11.002; -.
SQ   SEQUENCE   687 AA;  72017 MW;  6DD95714C350B3CD CRC64;

Alignment Scores:
Pred. No.:               1.07e-11      Length:         687
Score:                   281.50        Matches:        100
Percent Similarity:      34.45%        Conservative:   13
Best Local Similarity:   30.49%        Mismatches:     94
Query Match:             10.86%        Indels:         121
DB:                      10            Gaps:           15

SEQ12-SEQ4 (1-1390) x Q948Y7 (1-687)

Qy   1031 CGGGTTGAACTGTGCCTG---------------------TGCGTACCGATATCATTCAGC 993
             ||||||  ||||||                       |||   |||:::|||
Db    453 ArgValTyrLeuCysArgTyrAspValSerProAspGluCysProProLeuSerValLeu 472

Qy    992 GCCTGAATGCCCGCTTTCATACCGATACCGGTACCCACGGCGTTACCTAACTGCTGGTAG 933
                :::|||:::     |||   |||   |||   :::   |||
Db    473 GluThrLeuProSerAlaProProSerProProProProSerProPro------------ 488

Qy    932 TCCACCGGCCCGCTCAGG------TTTTGCAGCCCTTTGCCGCCCAGCGACGAACCGTCA 879
                   |||   |||            ||||||:::||||||:::      ||||||
Db    489 ---------ProProArgProProProProSerProValProProThrProProProSer 505

Qy    878 ---AGACCCGTGCCAGCA---TTACCGCCCTGACCACCTCCCAGTCCCCCGTTGCCAAGG 825
             ||||||   |||:::     ||||||   |||||||||   ||||||   ||||||
Db    506 ProArgProProProSerProArgProProAsnProProProArgProProSerProArg 525

Qy    824 AGCTGGCTCAGACCATTACCC------------ATCAGGCCCGACAGCGCATCAGTGACT 777
                ||||||   |||                 ||||||                 :::
Db    526 ProProProArgProProProArgProSerSerProArgProProProProAspProSer 545

Qy    776 CCTTTTTTATAGGCGTTCTGCTCGCCTTCGGTCGGCTGCTTGCCCCCAGAG--------- 726
          |||         :::   ||||||                  ||||||:::
Db    546 ProProProProSerProProSerProProThrSerProSerProProAspProAlaTrp 565

Qy    725 ---GAACTGCCCTGGGTGCCATCTTGCCCATCACCAAACAGGCTTTGCATTATCTCGCTG 669
                ||||||    |||      ||||||||||||
Db    566 AlaAsnLeuProThrSerProAspProProSerProAsn-------------------- 578

Qy    668 AACATCTTCAGCAGCTGCTGCATCGGGTCGCTGGAGTCTGAGGTGGAATCTGTGCCGGAG 609
                                                                   |||
Db    579 -------------------------------------------------ProProSer 581

Qy    608 GTGGAATCGTCGTTTTGGGACGTTGAGTTAATACCCAGCGCCTGGTCCAGCGGGGAATTT 549
                :::         |||         |||
Db    582 ProAspProProSerProAspProProSerAlaPro------------------------ 593

Qy    548 GTTGTTGAAGTGGTATTGTTGCCGCCTTTCGAGCCCAGCGTGTTCAGCGAACCGCCTAAC 489
                                ||||||         |||               |||||||||
Db    594 --------------------ProProSerProProProProSerProProProProAsn 606

Qy    488 ATATCGTTCAGCGCGTTCGACAGTCCTTCGCCCAGGCCACCTGAGCCACCCAAGCCATTA 429
                                 |||||||||   ||||||   ||||||   |||
Db    607 Pro---------------ProProProSerProProProProAsnProProProProSer 621

Qy    428 CCTAAGCCACCGCCTAAGCCACCGCCCATCAGCCCACCACCGCCCATCATGCTCATCATC 369
          |||   |||||||||   |||||||||   |||||||||||||||
Db    622 Pro---ProProProSerProProProProSerProProProPro--------------- 635

Qy    368 ATCATCATGCCGGTGAGTAAGCCAGCCAGCTGATTGACGGTATCATTTTGATTACCGCCG 309
                |||         |||                                   ||||||
Db    636 ---AsnProProProProSerPro----------------------------ProPro 644
```

```
Qy     308 CCCAGCCCCAGTGCAGAATTGCCACCCAACCCAGCATTCTGGCGACTGGTACCCAGCAAC 249
           |||||||||          ||||||   |||:::                  |||
Db     645 ProSerProArgProProThrProProProProSer---------------ProProPro 659

Qy     248 CCGTTATTTCCGCCCGCACCGCCGATAGAAATTGCATCGTTGACGCTCCCAGCCCACTT 189
           |||      ||||||   ||||||
Db     660 ProArgProProProArgProPro------------------------------------ 667

Qy     188 GTATTCAGACTCATAACGTATTCCTCTTATGAATAATTCAAACGTACCCTGCCGACCCAA 129
                                                              |||||||:::
Db     668 ---------------------------------------------------ProThrArg 670

Qy     128 CGGTCAAACGTGCCATGCCGAAGC 105
           ||||||      |||     |||
Db     671 ArgSerProProProThrSerSer 678

RESULT 8
Q8VKN7
ID   Q8VKN7      PRELIMINARY;      PRT;   598 AA.
AC   Q8VKN7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein MT0291.4.
GN   MT0291.4.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE006937; AAK44515.1; -.
DR   TIGR; MT0291; -.
KW   Hypothetical protein.
SQ   SEQUENCE   598 AA;  58523 MW;  75293DB1993B0ED6 CRC64;

Alignment Scores:
Pred. No.:              1.44e-11       Length:          598
Score:                  279.50         Matches:         121
Percent Similarity:     35.21%         Conservative:    35
Best Local Similarity:  27.31%         Mismatches:      150
Query Match:            10.78%         Indels:          137
DB:                     16             Gaps:            21

SEQ12-SEQ4 (1-1390) x Q8VKN7 (1-598)

Qy    1376 CCCAGCTTGCCAAGTGCCATATTGTTAATGGCATCACCGGCCATCATGGCATCAATACCC 1317
           ||||||   |||||||||    :::                |||      :::     :::|||
Db     163 ProSerProProSerAlaProValAlaProTrpProPro------LeuProProLeuPro 180

Qy    1316 AGCGAAGAACCACCGGCACCGCGTGCCTGCAGGTTGCCG---TTGCCGGTATCACCCGCC 1260
           :::        |||||||||||     :::        :::|||   :::|||::::::|||
Db     181 AsnAsnHisProProAlaProProSerAlaProValProGlyValProLeuAlaProLeu 200

Qy    1259 -----------------------------ATGGGCCTTTTGATCATGCCCTTGGCTTTG 1230
                                        :::|||   ||||||      |||    :::
Db     201 ProIleSerGlyArgProValArgAlaTrpValGlySerLeuIle---AlaLeuArgIle 219

Qy    1229 TTGAACTGCTCCATACTGGCTGGTGTCATTCCGTCGTCATCTGGCTTGCTCAGTGCTTTT 1170
                 |||    :::    :::||||||:::           :::|||   |||:::
Db     220 CysCysCysArgValCysSerGlyValLeu---------AlaGlyAlaLeuAsnProSer 236

Qy    1169 GCCCATGATTTGTCATCGGTTTTCACCTCCTGACCCGGGCCTTTCTGGTACTGCGGCTTG 1110
                     ||||||                  |||    |||           |||
Db     237 ArgPro------SerSerCysProProLysProProAlaProAlaValProAlaGlyAla 254

Qy    1109 CCAAACACCTCAGGATACTGGTCCATGAACTGACCGATTTCCTTCGCCATCGCCCGATCG 1050
           |||                           |||||||||      :::      :::
Db     255 ProValProProLeuProProLeuProProLeuProIleSerThrProLeuProProAla 274

Qy    1049 CCT--------------------TTATTGACGAAAGAACGGGTTGAACTGTGCCTGTGC 1011
           |||                     |||    |||
Db     275 ProProLeuProProLeuProAlaLeuProThrSerProGlyAlaProProAlaProPro 294

Qy    1010 GTACCGATATCATTCAGCGCCTGAATGCCCGCTTTCATACCGATACCGGTACCCACGGCG 951
           ||||||      :::                  ||||||     |||    |||    |||    |||
Db     295 ValProProAla---------------ProAlaLysAspProProProAlaProProAla 309

Qy     950 TTACCTAACTGCTGGTAGTCCACCGGCCCGCTCAGGTTT---------------TGCAGC 906
              |||                           ||||||
Db     310 ProPro------------------AlaProLeuSerArgProAlaPheProProAlaPro 323

Qy     905 CCTTTGCCGCCCAGCGACGAACCGTCAAGACCCGTGCCAGCATTACCGCCCTGACCA--- 849
           |||     ||||||:::     :::||||||    |||      |||          ||||||    |||
Db     324 ProAlaProProAlaSerLysProSerProProThrProProAlaProProGluProAsn 343

Qy     848 ------------------------CCTCCCAGTCCCCCGTTGCCAAGGAGCTGGCTCAGA 813
                                  ||||||:::||||||     |||
Db     344 AsnValProProSerProProIleProProAlaProProProProSerGlyLeuAspPro 363

Qy     812 CCATTACCCATCAGGCCCGACAGCGCATCAGTGACTCCTTTTTATAGGCGTTCTGCTCG 753
           |||||||||      |||    :::|||                  |||    |||              |||
Db     364 ProLeuProProAlaProProAlaAla---------ProArgLeuSerMetProAlaSer 380

Qy     752 CCTTCGGTCGGCTGCTTGCCCCCA-----------------GAGGAACTGCCCTGGGTG 711
           |||                  ||||||                    ||||||    |||
Db     381 ProProAlaProProPheProProThrLeuIleMetLeuValProProLeuProProVal 400

Qy     710 CCATCTTGCCCATCACCAAACAGGCTTTGCATTATCTCGCTGAACATCTTCAGCAGCTGC 651
           |||      |||:::||||||
Db     401 Pro------ProAlaProAsn--------------------------------------- 405

Qy     650 TGCATCGGGTCGCTGGAGTCTGAGGTGGAATCTGTGCCGGAGGTGGAATCGTCGTTTTGG 591
                                             |||   ||||||      |||
Db     406 -----------------------------SerProProGluProProSer---------- 412

Qy     590 GACGTTGAGTTAATACCCAGCGCCTGGTCCAGCGGGGAATTTGTTGTTGAAGTGGTATTG 531

Db     412 ------------------------------------------------------------ 412

Qy     530 TTGCCGCCTTTCGAGCCCAGCGTGTTCAGCGAACCGCCTAACATATCGTTCAGCGCGTTC 471
              ||||||         |||                  ||||||
Db     413 ---ProProAlaProProProLysMetProAsnProProGlyProProValProProAla 431

Qy     470 GACAGTCCTTCGCCCAGGCCACCTGAGCCACCCAAGCCA-----------------TTA 429
              :::        |||    ||||||:::||||||    |||
Db     432 ProAsnSerProProPheProProAspProProAlaProProValProAlaSerValAla 451

Qy     428 CCTAAGCCACCGCCTAAGCCACCGCCCATCAGCCCACCACCGCCCATCATGCTCATCATC 369
           |||        ||||||   ||||||          :::   ||||||
Db     452 ProProAlaProProThrProProSerAlaAsnSerProPro------------------ 465

Qy     368 ATCATCATGCCGGTGAGTAAGCCAGCCAGCTGATTGACGGTATCATTTTGATTACCGCCG 309
                       |||    :::   ||||||                              ||||||
Db     466 ------PheProProAlaProProAla--------------------------ProPro 474

Qy     308 CCCAGCCCCAGTGCAGAATTG------------CCACCCAACCCAGCATTCTGGCGACTG 261
           :::|||   |||                  ||||||   ||||||
Db     475 ValAlaProLysAlaAlaAlaAsnProProGlyProProThrProAlaAlaProAsnSer 494

Qy     260 GTACCCAGCAACCCGTTATTTCCGCCCGCACCGCCGATAGAAATTGCATCGTTGACGCT 201
           :::|||:::    |||      |||||||||||||||:::    :::          :::
Db     495 MetProAlaAlaProProAlaProProAlaProProValProValLeuAlaLeu---Pro 513

Qy     200 CCCAGCCCA 192
           |||:::|||
Db     514 ProAlaPro 516
```

```
RESULT 9
O53559
ID   O53559     PRELIMINARY;     PRT;   1489 AA.
AC   O53559;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PGRS-family protein.
GN   RV3514 OR MTV023.21.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; AL022022; CAA17751.1; -.
DR   TubercuList; Rv3514; -.
DR   InterPro; IPR000084; PE_region.
DR   InterPro; IPR002173; PfkB.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.
KW   Complete proteome.
SQ   SEQUENCE   1489 AA;  115754 MW;  6855CBA1C3CBAF3A CRC64;

Alignment Scores:
Pred. No.:              1.9e-10      Length:         1489
Score:                  264.50       Matches:        116
Percent Similarity:     32.13%       Conservative:   27
Best Local Similarity:  26.07%       Mismatches:     149
Query Match:            10.45%       Indels:         153
DB:                     16           Gaps:           16

SEQ12-SEQ4 (1-1390) x O53559 (1-1489)

Qy    181 CTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATT-------------------- 219
            |||||||||  ||||||   :::::|||
Db    955 LeuAsnThrAspGlyLeuSerSerAlaThrSerGlyThrGlyGlyThrGlyGlyThrGly 974

Qy    220 ---------------------------------------------------TCTATCGGC 228
                                                                  |||
Db    975 GlyLysGlyGlyThrGlyGlyAlaGlyAspAspSerAlaGlyGlyThrGlyGlyThrGly 994

Qy    229 GGTGCGGGCGGAAATAACGGGTTG--------------CTGGGTACCAGTCGCCAGAAT 273
          ||||||||||||||  |||                  |||  :::      |||
Db    995 GlyAlaGlyGlyAsnAlaGlyAlaGlyGlyLeuAlaAsnThrGlyGlyThrAlaGlyAsn 1014

Qy    274 GCTGGGTTGGGTGGCAATTCTGCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTC 333
          ||||||:::|||||||:::     |||  |||  |||  |||  |||:::
Db   1015 AlaGlyIleGlyGlyAspGlyGlyGlnGly---GlyAsnGlyGlyGlnGlyAspSerGly 1033

Qy    334 AATCAGCTG-----------------GCTGGCTTACTCACCGGCATGATGATGATGATG 375
          :::  |||                  ||||||  ·        |||
Db   1034 SerGlyLeuGlyGlyGlnProGlyPheAlaGlyGlyAlaGlyGlyLysGlyGlyAlaGly 1053

Qy    376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                |||   ||||||   |||  |||   ||||||  |||   |||
Db   1054 GlySerSerGlyAlaGlyGlyThrAsnGlySerGlyGlyAlaGlyGlyAlaGly---Gly 1072

Qy    436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
             ||||||:::||||||.  |||      ||||||    |||       |||
Db   1073 GlnGlyGlyAlaGlyGlyAlaGlyIleSerPheSerAsnGlySerAsnGlyGlyThrGly 1092

Qy    496 GGTTCG------CTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACA 549
          |||:::      :::        |||    ||||||||
Db   1093 GlyThrGlyGlyValGlyGlyThrGlyGlyAspGlyGlyAsn------------------ 1106

Qy    550 AATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACC 609
                                 |||    :::                |||
Db   1107 --------------------AlaGlyThrGlyAlaGlyAspProGlyLysGlyGlyThr 1119

Qy    610 TCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTC 669
             ||||||    |||    |||
Db   1120 GlyGlyThrGlyGlyThrGlyGlySerGly----------------------------- 1129

Qy    670 AGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACCCAGGGCAGTTCC--- 726
                                       |||     |||:::   |||::::::
Db   1130 -----------------------------GlyAlaGlyGlySerGlyGlyAlaAsnPhe 1139

Qy    727 TCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAGGAGTCACTGAT 786
          :::||||||     |||   |||                   ||||||
Db   1140 AsnGlyGlyThrGlyGlyThrGlyGlyThrGlyGlyThrGlyGlyLysGlyGlyMetGly 1159

Qy    787 GCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAACGGGGGACTGGGA 846
             :::::|||      |||    |||               ||||||   |||:::|||
Db   1160 GlyIleAlaGlyAspGlyGlyProGlyGlyAspGly------GlyAsnAlaGlyValGly 1177

Qy    847 ------------------------GGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGAC 882
                                  ||||||   ||||||      |||||||||
Db   1178 GlyLysGlyGlyThrAsnGlyAsnGlyGlySerGlyGlyThrGlyGlyThrGlyGlyAla 1197

Qy    883 GGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAG 942
          |||   :::   |||      ||||||   |||      |||
Db   1198 GlyGlyAsnAlaGlyAlaGlyGlyLeuAlaAsnThrGlyGlyThrAla------------ 1213

Qy    943 TTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGAT 1002
             |||||||||         |||||||||
Db   1214 ---GlyAsnAla---------GlyIleGly------------------------------ 1219

Qy   1003 ATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATG 1062
                                                          ||||||
Db   1220 -----------------------------------------------GlyAspGlyGlyGln 1224

Qy   1063 GCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTAC 1122
                ||||||    |||               :::   |||:::|||     :::
Db   1225 GlyGlyAsnGlyGlyGln---GlyAspSerGlySerGlyLeuGlyGlyGlnProGlyPhe 1243

Qy   1123 CAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGC 1182
             |||||||||
Db   1244 AlaGlyGlyProGlyGly------------------------------------------ 1249

Qy   1183 AAGCCAGATGACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGC 1242

Db   1249 ------------------------------------------------------------ 1249

Qy   1243 ATGATCAAAAGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCC 1302
             |||                  |||:::  |||   |||             ||||||
Db   1250 ------LysGlyGlyAlaGlyGlyAsnAlaGlyThrGlyGlyThrAsnGlySerGlyAla 1267

Qy   1303 GGTGGTTCTTCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCA 1362
          ||||||:::                  |||      ||||||   :::   :::|||   :::
Db   1268 GlyGlyAlaGlyGlyGlnGlyGlyAlaGlyGlyAlaGlyIleSerPheSerAsnGlySer 1287

Qy   1363 CTTGGCAAGCTGGGC 1377
             |||      |||
Db   1288 AsnGlyGlyThrGly 1292

RESULT 10
Q9SBM1
ID   Q9SBM1     PRELIMINARY;     PRT;   409 AA.
AC   Q9SBM1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hydroxyproline-rich glycoprotein DZ-HRGP precursor.
GN   HRGP GENE.
OS   Volvox carteri f. nagariensis.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Volvocaceae; Volvox.
OX   NCBI_TaxID=3068;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HK10;
RX   MEDLINE=20044763; PubMed=10574980;
RA   Ender F., Hallmann A., Amon P., Sumper M.;
RT   "Response to the sexual pheromone and wounding in the green alga
RT   Volvox: induction of an extracellular glycoprotein consisting almost
RT   exclusively of hydroxyproline.";
RL   J. Biol. Chem. 274:35023-35028(1999).
DR   EMBL; AJ242540; CAB62280.1; -.
DR   InterPro; IPR003882; Pistil_extensin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01582; KV33CHANNEL.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
KW   Signal.
FT   SIGNAL        1     17
FT   CHAIN        18    409       HYDROXYPROLINE-RICH GLYCOPROTEIN DZ-HRGP.
SQ   SEQUENCE   409 AA;  41547 MW;  CD0749C6AF02BD74 CRC64;

Alignment Scores:
Pred. No.:              1.96e-10      Length:          409
Score:                  263.00        Matches:         87
Percent Similarity:     37.11%        Conservative:    8
Best Local Similarity:  33.98%        Mismatches:      77
Query Match:            10.15%        Indels:          84
DB:                     10            Gaps:            12

SEQ12-SEQ4 (1-1390) x Q9SBM1 (1-409)

Qy     983 CCCGCTTTCATACCGATACCGGTACCCACGGCGTTACCTAACTGCTGGTAGTCCACCGGC 924
           |||:::       |||   |||   |||:::       |||             |||
Db     194 ProSerProProProSerProProProSerProProProArgProProProSerProAsn 213

Qy     923 CCGCTCAGGTTTTGCAGCCCTTTGCCGCCCAGCGACGAACCG------TCAAGACCCGTG 870
           |||             ||||||  ||||||||        |||        ||||||
Db     214 Pro---------ProSerProArgProProSerProSerProProSerProArgProPro 230

Qy     869 CCAGCATTACCGCCCTGACCACCTCCCAGTCCCCCGTTGCCAAGG---AGCTGGCTCAGA 813
           |||     ||||||   ||||||||||||||||||    ||||||
Db     231 ProArgArgProProPheProProProSerProProProProArgProProProProAla 250

Qy     812 CCATTACCCATCAGGCCCGACAGCGCATCAGTGACTCCTTTTTTATAGGCGTTCTGCTCG 753
           |||   |||   ||||||            |||        |||   |||
Db     251 ProProProProArgProProProProSerProProProProLeuProPro------Pro 268

Qy     752 CCTTCGGTCGGCTGCTTGCCCCCAGAGGAACTGCCCTGGGTGCCATCTTGCCCATCACCA 693
           ||||||                  ||||||      ||||||     ||||||   |||   |||
Db     269 ProSer------------ProProProProLeuProProProProSerProProProPro 284

Qy     692 AACAGGCTTTGCATTATCTCGCTGAACATCTTCAGCAGCTGCTGCATCGGGTCGCTGGAG 633

Db     284 ------------------------------------------------------------ 284

Qy     632 TCTGAGGTGGAATCTGTGCCGGAGGTGGAATCGTCGTTTTGGGACGTTGAGTTAATACCC 573
                                ||| |||                                         |||
Db     285 ------------SerProProPro-------------------------------Pro 289

Qy     572 AGCGCCTGGTCCAGCGGGGAATTTGTTGTTGAAGTGGTATTGTTGCCGCCTTTCGAGCCC 513
           |||                                    |||||||||
Db     290 SerPro---------------------------------LeuProProSerProArg 297

Qy     512 AGCGTGTTCAGCGAACCGCCTAACATATCGTTCAGCGCGTTCGACAGTCCTTCGCCCAGG 453
                                |||||||||     |||               |||:::|||
Db     298 ProProLysProSerProProAsnAsnProPheProArgProProProProAsnProPro 317

Qy     452 CCACCTGAGCCACCCAAGCCATTACCTAAGCCACCGCCTAAGCCACCGCCCATCAGCCCA 393
           ||||||  ||||||   |||   |||      ||||||:::|||||||||   ||||||
Db     318 ProProArgProProProProSerProAsn---ProProArgProProProProSerPro 336

Qy     392 CCACCGCCCATCATGCTCATCATCATCATCATGCCGGTGAGTAAGCCAGCCAGCTGATTG 333
           |||||||||                                         :::|||
Db     337 ProProPro--------------------------------ArgPro----------- 341

Qy     332 ACGGTATCATTTTGATTACCGCCGCCCAGCCCCAGTGCAGAATTGCCACCCAACCCAGCA 273
                             |||||||||||||||:::           ||||||   |||:::
Db     342 -----------------ProProProSerProAsnProProArgProProProProSer 355

Qy     272 TTCTGGCGACTGGTACCCAGCAACCCGTTATTTCCGCCCGCACCGCCG 225
                        |||      |||      ||||||:::||||||
Db     356 ---------ProSerProProLysProProProProProSerProPro 368

RESULT 11
Q98LN6
ID   Q98LN6      PRELIMINARY;      PRT;  3659 AA.
AC   Q98LN6;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein mll0950.
GN   MLL0950.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP002996; BAB48427.1; -.
DR   InterPro; IPR002173; PfkB.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   3659 AA;  337498 MW;  7B7BC7EC5C0F3816 CRC64;

Alignment Scores:
Pred. No.:              4.19e-10      Length:          3659
Score:                  260.50        Matches:         133
Percent Similarity:     37.90%        Conservative:    55
Best Local Similarity:  26.81%        Mismatches:      186
Query Match:            10.29%        Indels:          123
DB:                     16            Gaps:            21

SEQ12-SEQ4 (1-1390) x Q98LN6 (1-3659)

Qy     178 AGTCTGAATACAAGTGGGCTGGGAGCG---TCAACGATGCAAATTTCTATCGGCGGTGCG 234
            ||||||   |||   |||   |||:::                   ||||||||||||
Db     658 SerLeuThrThrTyrGlyGlnGlySerArgGlyIlePheAlaGlnSerIleGlyGlyThr 677

Qy     235 GGCGGAAATAACGGG----------------------TTGCTGGGTACCAGTCGCCAG 270
            |||||||||   |||                                        ||||||
Db     678 GlyGlyAsnGlyGlyAlaAlaIleAlaValAlaValAsnAlaThrGlyThrPheAlaAla 697

Qy     271 AATGCTGGGTTGGGTGGCAATTCTGCACTGGGGCTGGGCGGCGGTAATCAAAATGATACC 330
                  |||:::|||||||||         |||   ||||||   ::::::      |||
Db     698 AlaIleGlyIleGlyGlyAsnGly------GlyAlaGlyGlyAlaSerLysLeuValThr 715

Qy     331 GTCAATCAGCTGGCTGGCTTACTCACC--------------GGCATGATGATGATGATG 375
            |||:::   |||:::   :::   |||                        |||   ::::::
Db     716 ValSerSerLeuSerAspIleSerThrAlaGlyAsnAsnAlaGlyGlyIleValAlaGln 735
```

```
Qy    376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
          |||:::   |||||||||||        |||      :::||||||   :::|||
Db    736 SerValGlyGlyGlyGlyGlyAsnGlyGlyTyrAlaIleGlyGlyAlaIleGlyGlyAla 755

Qy    436 TTGGGT---------GGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGAT 486
          :::|||         ||||||   |||   |||               |||     |||
Db    756 ValGlyIleAlaValAsnIleGlyGlyLysGlyAlaGly---GlyGlyAlaGlyAlaAsp 774

Qy    487 ATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACA 546
          ::::::   ||||||      ::::::   :::||||||   |||:::|||
Db    775 ValIleValGlySerSerGlySerIleHisThrLysGlyAsnAsnSerThrGlyIleLeu 794

Qy    547 ACAAATTCC-----------------------------CCGCTGGACCAGGCGCTGGGT 576
             |||                                :::       |||||||||
Db    795 AlaGlnSerValGlyGlyGlyGlyGlyAsnGlyGlyPheThrIleSerAlaAlaLeuGly 814

Qy    577 ATTAACTCAACGTCCCAAAAC-----------------GACGATTCCACCTCCGGCACA 618
             :::   |||                          |||   |||   :::|||
Db    815 GlyGlyAlaAlaSerValGlyLeuGlyGlyAlaGlyAlaAspGlySerAsnAlaGly--- 833

Qy    619 GATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATA 678
             :::   :::   |||:::|||               :::
Db    834 ---AsnValThrValSerAlaAspGlyAlaGlyHisThrValAlaThrThrTyrThrGly 852

Qy    679 ATGCAAAGCCTGTTTGGTGATGGGCAAGAT------GGCACCCAGGGCAGTTCCTCTGGG 732
             :::|||   :::|||::::::      |||   |||      |||   |||
Db    853 ThrTrpAsnLeuValThrGluGlyLysAsnAlaValGlyIleGlnAlaGluSerIleGly 872

Qy    733 GGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAGGAGTCACTGATGCGCTG 792
          |||               |||      :::         |||
Db    873 GlyGlyGlyGlyAsnGlyGlyPheSerGlySerLeuAlaAlaGly--------------- 887

Qy    793 TCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAACGGGGGACTGGGAGGTGGT 852
             ||||||:::|||   |||:::|||      ||||||   |||   |||||||||
Db    888 ---GlyLeuValGlyValGlyValSer------LeuGlyGlyThrGlyGlyGlyGlyGly 904

Qy    853 CAGGGCGGTAATGCT------------------------------------------- 867
             |||   |||
Db    905 AspAlaGlyThrAlaThrValAsnAsnGlyLysLysValGlyAsnThrValThrHisAsn 924

Qy    868 ------------------------GGCACGGGTCTTGACGGTTCGTCGCTGGGCGGCAAA 903
                                  |||||||:::      |||:::|||||||
Db    925 AsnIleLeuThrIleGluAspAsnSerThrGlyIleLeuAlaGlnSerIleGlyGlyGly 944

Qy    904 GGG---------------CTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGT 948
          |||                        ||||||      |||:::      |||
Db    945 GlyGlyAsnGlyGlyPheAlaValThrLeuSerValSerGlySerTyrGluGlyAlaGly 964

Qy    949 AACGCCGTGGGTACC--------------------------GGTATCGGTATGAAAGCG 981
             |||                                   |||:::|||
Db    965 GlyAlaAlaAlaValSerValGlyGlySerGlyAlaThrGlyGlyValGlyLysAspVal 984

Qy    982 GGCATTCAGGCGCTGAATGATATCGGTACGCACAG-GCACAGTTCAACCCGTTCTTTCGT 1040
             :::   :::      :::|||   |||:::      |||   :::
Db    985 PheValThrSerTyrGlyAsnIleAlaThrTyrGlyLysGlnSerAspGlyIleLeuAla 1004

Qy   1041 CAATAAAGGCGATCGGGCGATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCTGA 1100
          |||               |||   |||   |||      ||||||   |||      :::
Db   1005 GlnSerIleGlyGlyGlyGlyGlyAsnGlyGlyPheSerValAlaGlyThrPheThrThr 1024

Qy   1101 GGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGACAA 1160
          |||         |||:::|||            |||||||||||            |||
Db   1025 GlyAlaLeuGlyAlaSerValAlaVal---GlyGlySerGlyGlySerGly------Gln 1041

Qy   1161 ATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATGGA 1220
                |||:::   |||                                :::   |||
Db   1042 SerAlaGlyGluValThr---------------------------ValThrSerAlaGly 1052

Qy   1221 GCAGTTCAA------------CAAAGCCAAGGGCATGATCAAAAGGCCCATGGCGGGTGA 1268
             :::|||              ||||||   |||:::::   :::   :::   |||
Db   1053 AsnIleGlnThrHisGlyAspGlnSer-IleGlyIleMetAlaGlnSerValGlyGlyGl 1072

Qy   1269 TACCGGCAACGGCAACCTGCAGGCACGCGGTGCC------------------------- 1302
                |||||||||         ||||||
Db   1072 yGlyGlyAsnGlyGlyPheAlaGlyAlaGlyAlaIleThrLeuGlnGlyValSerAlaAl 1092

Qy   1303 ---------------------------GGTGGTTCTTCGCTGGGTATTGATGCCATGAT 1334
                                     ||||||||||:::         :::
Db   1092 aValGlyLeuGlyGlySerGlyAlaGlyGlyGlySerAlaLysThrValArgValThrSe 1112

Qy   1335 GGCCGGTGAT------GCCATTAACAATATGGCACTTGGCAAGCTG 1374
                ||||||            :::   |||:::|||   |||
Db   1112 rThrGlyAspIleThrThrTyrGlyAspGlnAlaIleGlyIleLeu 1127
```

```
RESULT 12
O53215
ID   O53215      PRELIMINARY;      PRT;  1665 AA.
AC   O53215;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PGRS-family (PE_PGRS family protein).
GN   RV2490C OR MT2564 OR MTV008.46C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021246; CAA16067.1; ALT_INIT.
DR   EMBL; AE007093; AAK46868.1; -.
DR   TIGR; MT2564; -.
DR   TubercuList; Rv2490c; -.
DR   InterPro; IPR000084; PE_region.
DR   InterPro; IPR002173; PfkB.
DR   InterPro; IPR000228; RNA3'_term_cycl.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
DR   PROSITE; PS01287; RTC; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   1665 AA;  133700 MW;  2F300BD7766994E9 CRC64;

Alignment Scores:
Pred. No.:               4.7e-10       Length:         1665
Score:                   259.00        Matches:        119
Percent Similarity:      33.41%        Conservative:   22
Best Local Similarity:   28.20%        Mismatches:     144
Query Match:             10.23%        Indels:         137
DB:                      16            Gaps:           19
```

```
SEQ12-SEQ4 (1-1390) x O53215 (1-1665)

Qy   193 GGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAATAACGGGTTG 252
Db   147 GlyAsnGlyAlaAlaGlyAlaValGlyGlnValGlyValGlyAlaGlyGlyAlaAlaGlyLeu 166

Qy   253 CTGGGTACCAGTCGCCAGAAATGCTGGGTTGGGTGGCAATTCTGCACTGGGGCTGGGCGGC 312
Db   167 PheGlyIle-----------GlyGlyAlaGlyGlyAlaGlyGlyAlaGlyAlaProGly 182

Qy   313 GGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATG 372
Db   183 GlyThr---GlyGlyThrGlyGlyTrpLeuAlaGly---------- 193

Qy   373 ATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAAT 432
Db   194 GlyGlyGlyValGlyGlyMetGlyGlyAlaGlyGlyGlyAlaGlyGlyAlaGlyGlyAsn 213

Qy   433 ----------GGCTTGGGTGGCTCAGGTGGCCTGGCCGAAGGACTGTCG 471
Db   214 AlaGlyLeuPheGlyAsnGlyGlyAlaGlyGlyAlaGlyGlyAlaGlyGlyAlaGly 233

Qy   472 AACCGCCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAGGCGGCAAC 531
Db   234 GlyAla----------GlyGlyAsnAlaGlyTrpPheGlyHisGlyGlyAlaGly 248

Qy   532 AATACCACTTCAACAACAAATTCCCCGCCTGCACCAGCGCCTGGGTATTAACTCAACGTCC 591
Db   249 GlyValGlyGlyVal----------GlyAlaAlaGlyAlaAsnGlyAlaThr 262

Qy   592 CAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAG 651
Db   263 ProGlyGlnAspGlyAlaAlaGlyValAlaGlySerAspAspGlyAlaGly--------- 279

Qy   652 CAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGC 711
Db   280 ----------GlyAspGlyLeuAlaGly 285

Qy   712 ACCCAGGGCAGTTCCTCTGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
Db   286 SerAspGlyGlyAspGlyGlyAla---------- 293

Qy   772 AAAGGAGTCACTGATGCGCTGTCGGCCTGATGGGTAATGGT------CTGAGCCAGCTC 825
Db   294 ----------GlyGlyValGlyGlyAsnGlyGlyArgGlyGlyTrpLeu 306

Qy   826 CTTGGCAACGGGGCACTGGGA------GGTGGTCAGGCCGGTAATGCTGGCACGGGTCTT 879
Db   307 LeuGlyAsnGlyGlyAlaGlyGlyValGlyGlyValGlyGlyAlaGlyGlyAlaGlyAla 326

Qy   880 GACGGTTCGTCGCTGGCGGCAAAGGGCTCAAAACCTGAGCGGGCCGGTGGACTACCAG 939
Db   327 AlaGlyGlyAla---GlyGlyAlaGlyAlaThrGlyIleAsnGlyProAlaGlyIleSer 345

Qy   940 CAGTTAGGTAACGCCGTGGGTACC---GGTATCGGTATGAAAGCGGGCATTCAGGCGCTG 996
Db   346 AlaAlaGlyGlyAspGlyGlyAlaGlyGlyAsnGlyGlyAlaGlyGlyAsnGlyGlyVal 365

Qy   997 AATGATATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGG 1056
Db   366 GlyGlyAlaGly-GlyAlaGlyGlySerAlaGlyLeuLeuGlyTyrVal-GlyArgAlaG 385

Qy  1057 GCGAT----------GGCGAAGGAAATCGGT 1077
Db   385 lyAspGlyGlyAlaGlyGlyGlyGlyLeuGlyGlyAlaProGlyAspGlyGlyAlaG 405

Qy  1078 CAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGT 1137
Db   405 lyGlyAsnGly----------GlySerTrpLeuAlaAla----------GlyA 416

Qy  1138 CAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGAC 1197
Db   416 spGlyGlyAlaGly---------- 420
```

```
Qy   1198 GGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGGCCC 1257
Db    421 ----------GlyH 422

Qy   1258 ATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTQCCGGTGGTTCTTCGCTG 1317
Db    422 isGly-GlyAspProGlyLeuGlyGly------AlaGlyGlyAlaGlyGlyAlaSerGly 439

Qy   1318 GGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTGGGC 1377
Db    440 GlyAlaGlyAla---ArgAlaGly------AlaAsnGlyLeuAlaAlaGlyAsnAspGly 456

RESULT 13
Q8VJ19
ID   Q8VJ19       PRELIMINARY;      PRT;   628 AA.
AC   Q8VJ19;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PE_PGRS family protein.
GN   MT3476.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE007154; AAK47814.1; -.
DR   TIGR; MT3476; -.
DR   InterPro; IPR000104; Antifreeze_1.
DR   InterPro; IPR001343; Hemlysn_Ca_bind.
DR   InterPro; IPR000084; PE_region.
DR   InterPro; IPR002173; PfkB.
DR   Pfam; PF00934; PE; 1.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00313; CABNDNGRPT.
DR   ProDom; PD001223; PE_region; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
SQ   SEQUENCE   628 AA;  53828 MW;  9F62B1E7CC2C567E CRC64;

Alignment Scores:
Pred. No.:              4.72e-10        Length:         628
Score:                  258.00          Matches:        124
Percent Similarity:     36.34%          Conservative:   17
Best Local Similarity:  31.96%          Mismatches:     148
Query Match:            10.19%          Indels:         99
DB:                     16              Gaps:           21

SEQ12-SEQ4 (1-1390) x Q8VJ19 (1-628)

Qy     226 GGCGGTGCGGGCGGA------AATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGG 279
Db     320 GlyGlyAlaGlyGlyAspAlaAsnGlyAsnProAlaAsnThrSerIleAlaAsnAlaGly 339

Qy     280 TTGGGTGGCAATTCTGCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAG 339
Db     340 AlaGlyGlyAsnGlyAlaAlaGly---GlyAspGlyGlyAlaAsnGlyGlyAlaGlyGly 358

Qy     340 CTGGCTGGCTTACTCACCGGCATGATGATGATGATGACATGATGGGCGGTGGTGGGCTG 399
Db     359 AlaGlyGlyGlnAlaAlaSerAlaGlySerGlyValGlyGlyAspGlyGlyAsnGlyGly 378

Qy     400 ATGGGCGGT----------GGCTTAGGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCA 447
```

```
               ||||||           |||      ||||||   |||   |||   ||||||:::
Db     379 AlaGlyGlyThrGlyThrAsnGlyHisAlaGlyGlyAlaGly---GlyAlaGlyGlyAla 397

Qy     448 GGTGGCCTGGGCGAAGGACTG---TCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTG 504
           ||||||   |||      |||      |||      |||      |||||||:::
Db     398 GlyGlyArgGlyGlyTrpLeuValGlyAsnGlyGlyAsnGlyGlyAsnGlyGlyAsn--- 416

Qy     505 AACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCCCCGCTGGAC 564
                    |||:::   |||||||||               |||
Db     417 ---------GlyAlaAlaGlyGlyAsnGlyAlaIleGlyGlyThrGlyGly--------- 430

Qy     565 CAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCC 624
              |||   ||||:::            :::|||      :::|||      ||||||
Db     431 ---AlaGlyGlyVal---ProAlaAsnGlnGlyGlyAsnSerAlaLeuGlyThr------ 446

Qy     625 ACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAA 684
                                                                     |||
Db     447 ---------------------------------------------------------Gln 447

Qy     685 AGCCTGTTTGGTGATGGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCG 744
              :::   |||||||||   ||||||      |||   :::   |||
Db     448 ProValGlyGlyAspGlyGlyAspGlyGlyAsnGlyGlyThrGlyGly------------ 463

Qy     745 ACCGAAGGCGAGCAGAACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGC---CTG 801
           |||   |||                     |||         ||||||   |||
Db     464 ThrGlyGlyArgGlyGlyAspGlyGlySerGlyGlyAlaGlyGlyAlaSerGlyTrpLeu 483

Qy     802 ATGGGTAATGGTCTGAGCCAGCTCCTTGGCAACGGGGGACTGGGA------GGTGGTCAG 855
           |||||||||||   :::         |||||||||||   |||         ||||||
Db     484 MetGlyAsnGlyGlyAsnGly------GlyAsnGlyGlyThrGlyGlySerGlyGlyVal 501

Qy     856 GGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTGGCGGCAAAGGGCTGCAAAAC 915
           ||||||||   |||   |||   ||||||:::         ||||||
Db     502 GlyGlyAsnGlyGlyIleGlyGlyAspGlyAlaGly---GlyGlyAsnAlaThrSerThr 520

Qy     916 CTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATG 975
              |||   |||   |||            |||         |||   |||   |||  :
Db     521 SerSerIleProPheAspAlaHis------GlyGlyAsnGlyGlyAlaGly---Gly--A 537

Qy     976 AAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCACAGGCACAGTTCAACCCGTTCT 1035
           :::::||||||   |||
Db     537 spAlaGlyHisGlyGlyThr---------------------------------------- 543

Qy    1036 TTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTAT 1095
                                |||||||||   ||||:::                |||
Db     544 -----------------GlyGlyAspGly---GlyAspGlyGlyHisAlaGlyThrGlyG 557

Qy    1096 CCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGAT 1155
                 |||         |||            |||:::||||||
Db     557 lyArgGlyGlyLeuLeuAlaGlyGlnHisAlaAsnSerGlyAsnGlyGly---------- 573

Qy    1156 GACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGT 1215
                       |||      |||      |||                  ||||||   |||
Db     574 ----------GlyGlyGlyThrGlyGlyAlaGlyGlyThrHisGly-ThrProGlySer 589

Qy    1216 ATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGGCCCATGGCGGGTGATACCGGC 1275
                                                        ||||||   ||||||
Db     590 GlyAsn---------------------------------------AlaGlyGlyThrGly 596

Qy    1276 AACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCGCTGGGTATTGATGCCATGATG 1335
              ||||||      :::               |||      |||   |||            :::
Db     597 ThrGlyAsnAlaAspSerThrAsnGlyGlyProGlySerAspGly-----------Leu 612

Qy    1336 GCCGGTGATGCCATTAAC 1353
              |||||||||   |||
Db     613 GlyGlyAspAlaPheAsn 618
```

RESULT 14
Q92UU8

```
ID   Q92UU8       PRELIMINARY;      PRT;   2174 AA.
AC   Q92UU8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical glycine-rich protein SMb21548.
GN   RB0989 OR SMB21548.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OG   Plasmid pSymB (megaplasmid 2).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396508; PubMed=11481431;
RA   Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA   Vorhoelter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA   Golding B., Puehler A.;
RT   "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT   fixing endosymbiont Sinorhizobium meliloti.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR   EMBL; AL603645; CAC49389.1; -.
DR   InterPro; IPR002016; Peroxidase.
DR   InterPro; IPR002173; PfkB.
DR   PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.
KW   Plasmid; Hypothetical protein; Complete proteome.
SQ   SEQUENCE   2174 AA;  203314 MW;  008EB68297B44182 CRC64;

Alignment Scores:
Pred. No.:                6.26e-10      Length:          2174
Score:                    257.50        Matches:         125
Percent Similarity:       38.58%        Conservative:    49
Best Local Similarity:    27.72%        Mismatches:      163
Query Match:              10.17%        Indels:          114
DB:                       16            Gaps:            25

SEQ12-SEQ4 (1-1390) x Q92UU8 (1-2174)

Qy     133 GTCGGCAGGGTACGTTTGAATTATTCATAAGAGGAATACGTTATGAGTCTGAATACAAGT 192
           ||||||   :::::::::|||                        |||     |||:::   |||
Db     343 ValGlyAlaIleGlnIleAsnVal-----------TyrGlyGlySerIleGluThrAsp 358

Qy     193 GGGCTGGGAGCG---TCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAATAACGGG 249
           |||:::   |||            :::      |||||||||||   |||||||||   |||
Db     359 GlyIleGlnAlaTyrGlyLeuLeuGlyGlySerIleGlyGlyGlnGlyGlyAsnGlyGly 378

Qy     250 -----------TTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTGCA 297
                       ::::::|||||||||                  |||   |||||||||:::
Db     379 AspAspThrAlaValValGlyThrSer---GlyGlyGlyGlyPheGlyGlyAsnAlaGly 397

Qy     298 CTGGGGCTGGGCGGC-----------GGTAATCAAAATGATACCGTCAATCAGCTGGCT 345
              |||:::|||                |||::::::      |||             :::
Db     398 ---GlyValGlyAlaTyrValThrThrGlySerLysIleThrThrThrGlyAspPheSer 416

Qy     346 GGCTTACTCACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGC 405
                 :::|||                  :::      :::|||||||||||      |||
Db     417 AlaAlaValThr--------------LeuHisSerIleGlyGlyGlyGlyGlyThrGly 431

Qy     406 GGTGGC-----------TTAGGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGC 453
           |||                |||||||||||||   ||||||||      |||      |||
Db     432 GlyAspPheThrAspValLeuGlyGlyGlyAlaGlyAsnGlyGlyAsnGlyGlyAsnGly 451

Qy     454 CTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTG 513
                 ::::::||||||                  ||||||:::::::::|||
Db     452 AsnThrAlaThrIleAsnAsnAla--------------GlyGlyThrIleSerThrSer 466

Qy     514 GGC---------------------------TCGAAAGGCGGCAACAATACCACTTCAACA 546
           |||                                |||   ||||||:::      |||
Db     467 GlyGlyHisSerTyrGlyIleLeuValGlnSerIleGlyGlySerGlyGlyThrGlyGly 486

Qy     547 ACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCC 606
                 |||         ||||||   :::                        :::
Db     487 IleAlaGluGlyLeuThrLeuGluLeuGlyGlyAspGly---------GlyGluGlyGly 503

Qy     607 ACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATG 666
```

```
             ::::::|||        |||       ::::::
Db      504 SerAlaGlyAlaAlaSerValGlnAsnThrGlyAlaIleThrThrAspGlyTyrSerAla 523

Qy      667 TTCAGCGAGATAATGCAAAGCCTGTTT---GGTGATGGGCAAGATGGCACCCAGGGC--- 720
                :::   ||||||:::           |||    |||      ||||||   |||
Db      524 HisGlyIleLeuAlaGlnSerIleSerGlyGlyGlyGlyAlaAlaGlyThrAlaGlyGly 543

Qy      721 -----------AGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTAT 768
                              :::||||||             |||
Db      544 IleLeuSerIleGlyGlyThrGlyGlyAsnAlaAsnTyrAlaGlyAsnAlaGluValGln 563

Qy      769 AAAAAAGGAGTCACT-----------GATGCGCTGTCGGGCCTGATGGGTAATGGTCTG 816
                   |||                  ||||||      |||::::::         :::
Db      564 SerSerGlyAlaIleSerThrGlyGlyAspAlaAlaIleGlyIleLeuAlaGlnSerIle 583

Qy      817 AGCCAGCTCCTTGGCAACGGGGGACTGGGAGGTGGTCAG-----------------GGC 858
                            |||    ||||||   |||||||||                |||
Db      584 Gly---------GlyGlyGlyGlySerGlyGlyGlyAlaLysGlyIleAlaThrValGly 600

Qy      859 GGTAATGCTGGCACGGGT-------------------------CTTGACGGTTCGTCG 891
            |||:::   |||   |||                 .       |||:::|||   |||
Db      601 GlySerGlyGlyAlaGlyGlyTyrGlyThrTyrAlaSerValTrpLeuAsnGlyGlySer 620

Qy      892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
            :::      :::|||         ::::::      :::      |||   :::|||
Db      621 IleAlaThrArgGly------GluMetAlaHisGlyIleMetAlaGlnSerIleGlyGly 638

Qy      952 GCCGTGGGTACCGGT--------ATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGAT 1002
                   |||   |||           |||   |||       ||||||   ||||||
Db      639 GlyGlyGlyAsnGlyGlyAsnValIleAspMetSerValGlyIleProAlaLeuGly--- 657

Qy     1003 ATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATG 1062
            :::|||                                     |||   |||       |
Db      658 ValGly-------------------------------------Gly--SerAlaThrG 664

Qy     1063 GCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTAC 1122
            ||   |||:::                         |||   |||
Db      664 lyGlyGlySer--------------------------GlyGlyTrpValCysValThrA 675

Qy     1123 CAGAAAGGCCCGGG---------TCAGGAGGTGAAAACCGATGACAAATCATGGCAAAA 1173
               |||   ||||||        |||   |||         |||.  :::      :
Db      675 snGluAsnProGlyGlyAspCysSerThrGlySerGlyAlaProGlnAlaValAlaIleA 695

Qy     1174 GCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAA 1233
            ::|||      |||                                  |||   :::   ||||
Db      695 laThrAlaGlyAla--------------------GlyAlaValGlyIleLeuAlaGlnS 708

Qy     1234 GC-----------CAAGGCATGATCAAAAGGCCCATGCGGGTGATACCGGCAACGGC 1281
            ||                   |||:::         |||   |||  ||||||   |||   |||
Db      708 erIleGlyGlyGlyGlyGlyAsnGlyGlyAsnAlaThrGly-GlyAspAlaGlyPheGly 727

Qy     1282 AACCTGCAGGCACGCGGTGCCGGTGGT 1308
            :::    |||      |||   ||||||
Db      728 SerPheGlnIleGlyGlyGlyGlyGly 736
```

RESULT 15
O53552
ID   O53552   PRELIMINARY;   PRT;   1381 AA.
AC   O53552;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PGRS-family protein.
GN   RV3507 OR MTV023.14.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; AL022022; CAA17744.1; -.
DR   TubercuList; Rv3507; -.
DR   InterPro; IPR000084; PE_region.
DR   InterPro; IPR002173; PfkB.
DR   Pfam; PF00934; PE; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.
KW   Complete proteome.
SQ   SEQUENCE   1381 AA; 110624 MW; CA09676BD07F6482 CRC64;

Alignment Scores:

Pred. No.:	8.04e-10	Length:	1381
Score:	255.50	Matches:	128
Percent Similarity:	35.04%	Conservative:	29
Best Local Similarity:	28.57%	Mismatches:	187
Query Match:	10.09%	Indels:	105
DB:	16	Gaps:	21

SEQ12-SEQ4 (1-1390) x O53552 (1-1381)

```
Qy      184 AATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAAT 243
            |||      |||   |||                    ||||||||||||||||||
Db      479 AsnProGlyGlyLysGlyGlyAspGly---------GlyIleGlyGlyAlaGlyGlyAla 495

Qy      244 AACGGGTTGCTGGGTACCAGTCGC-----------CAGAATGCTGGGTTGGGTGGCAAT 291
               |||      ||||||                   |||   |||   ||||||:::
Db      496 GlyGlyAlaAlaGlyThrGlyAsnGlyGlyHisAlaGlyAsnThrGlyAspGlyGlyAsp 515

Qy      292 TCTGCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTA 351
                     |||      |||||||||               ||||||
Db      516 GlyGlyThrGlyGlyAsnGlyGlyAsnGlyThrGlyGlyValAsnGlyAlaAspAsnThr 535

Qy      352 CTCACCGGCATGATGATGATGATGAGCATGATGGGCGGT-----------GGTGGGCTG 399
            |||                           ||||||          ||||||
Db      536 LeuAsn--------------ProAspThrProGlyGlyAlaGlyGluProGlyGlyAla 550

Qy      400 ATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGC 459
               |||   |||      ||||||          |||   ||||||:::||||||   |||
Db      551 GlyGlyAlaGlyGlyAlaGlyGlyAlaAlaGlyGlyProGlyGlyThrGlyGlyThrGly 570

Qy      460 GAAGGACTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCG 519
                        |||      |||         ||||||:::            |||
Db      571 GlyAsnGlyGlyAsnGlyGlyAsnGlyGlyAsnGlyGlyAsnGlyGlyAsn---GlyGly 589

Qy      520 AAAGGCGGCAACAATACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATT 579
               |||||||||         ::::::||||||:::|||               :::|||
Db      590 AsnGlyGlyAsnAlaGlyAsnAsnSerThrAsnAlaPro-----------ValGlyGly 605

Qy      580 AACTCAACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGC 639
                     :::      |||               |||         ::::::
Db      606 GluGlyGlyAlaGlyGlyAspGlyGlyAlaGlyGlyAlaGlyGlyAlaAlaAsnGlyGly 625

Qy      640 GACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGAT 699
                                         :::         |||         |||
Db      626 ----------------------------ThrAlaGlySerGlnGlyThrGlyGlyVal 635

Qy      700 GGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAG 759
            |||   ||||||      |||   :::   ||||||   :::         |||
Db      636 GlyGlyAspGlyGlyAlaGlyGlyAsnGlyGlyGlyGlyLys------AlaGlyThrGly 653

Qy      760 AACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGC 819
            |||:::         ||||||   |||      :::|||      |||
Db      654 AsnSerGlyAsnPheGlyVal---AspGlyGluAlaGlyPheSerGlyGlyAla------ 670
```

```
Qy    820 CAGCTCCTTGGCAACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTT 879
          ||||||||||||:::||||||     |||   |||   ||||||||
Db    671 ------GlyGlyAsnGlyGlyValGlyGlyAlaAlaGlyAlaAsnGlyGlyThrGlyGly 688

Qy    880 GACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAG 939
          |||   :::   |||   |||       :::   |||
Db    689 SerGlyGlyAsnGlyGlyAspGlyGlyAlaGlyGlyIleGlyGlyAlaGly--------- 705

Qy    940 CAGTTAGGTAACGCCGTG----------------------------GGTACCGGTATC--- 969
          ||||||   :::                               |||||||
Db    706 ------GlyAsnGlyIleProGlyThrGlyThrGluProAlaGlyGlyThrGlyAlaLys 723

Qy    970 GGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCACAGGCACAGTTCAACC 1029
          |||       |||                 |||   :::       :::
Db    724 GlyGlyAspGlyGlyAspGlyGlyAlaGlyGlyAlaGlyGlyAsnAlaGlyGlyAlaGly 743

Qy   1030 CGTTCTTTCGTCAATAAAGGCGA-------TCGGGCGATGGCGAAGGAAATCGGTCAGTT 1082
               |||   |||       :::|||       ||||||       :::
Db    744 GlyGlnGlyGlyAsnAlaGlyGlnGlyGlyAlaGlyGlyAlaGlyGlyAsnAlaValIle 763

Qy   1083 CATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGTCAGGA 1142
          |||       ||||||   :::|||       :::       |||||||||
Db    764 ProGlyAsp---------GlyValGlyLysAlaProHisGlyAspAlaGlyGlySerGly 780

Qy   1143 GGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGACGGAAT 1202
          |||:::       ||||||       :::
Db    781 GlyAspGly--------------GlyLysGlyGlyGlnGlyGly---------SerGly 792

Qy   1203 GACACCAGCCAGTATGGAGCA----------------------GTTCAACAAAGCCAA 1238
          |||       ||||||                        |||
Db    793 GlyThrGlySerGlyAlaProIleGlyGlyGlyAlaGlySerGlyThrGlyGlySerGly 812

Qy   1239 GGGCATGATCAAAAG------------------------------------------- 1253
          ||||||     |||
Db    813 GlyHisAlaGlyLysGlyGlyAlaGlyGlyIleGlyAlaGlnGlyThrThrIleThrVal 832

Qy   1254 -----------GCCCATGGCGGGTGATACCGGCAAC---GGCAACCTGCAGGCACGCGG 1298
                    ::: |||||||||   ||||||   ||||||      |||   ||
Db    833 ProGlyAsnGlyGlyAsn-AlaGlyAspGlyGlyAsnGlyGlyAsnAlaGlyAlaGlyGl 852

Qy   1299 TGCCGGTGGTTCTTCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATAT 1358
          |  ||||||||                          ||||||      |||
Db    852 yAsnGlyGlySer----------------------------GlyAspPheGlyGlyAsnTh 863

Qy   1359 GGCACTTGGCAAGCTGGGCGCG 1380
               |||   |||:::
Db    863 rThrSerGlyAlaSerGlySer 870

Search completed: January 14, 2003, 17:11:17
Job time : 67.5 secs
```

```
                    GenCore version 5.1.3
          Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:         January 14, 2003, 17:00:07 ; Search time 15 Seconds
                                             (without alignments)
                                             7686.946 Million cell updates/sec

Title:          SEQ12-SEQ4
Perfect score:  2532
Sequence:       1 atgggatttttctcttttc..........gctgggcgcggcttaagctt 1390

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters:       225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/KUBELIK693/runat_14012003_091818_7067/app_query.fasta_1.1543
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=KUBELIK693_@CGN_1_1_15_@runat_14012003_091818_7067 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      SwissProt_40:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2050	81.0	403	1	HRPN_ERWAM	Q01099 erwinia amy
	2	828.5	32.7	356	1	HRPN_ERWCA	Q47279 erwinia car
	3	718.5	28.4	340	1	HRPN_ERWCH	Q47278 erwinia chr
c	4	316.5	12.2	555	1	GP1_CHLRE	Q9fpq6 chlamydomon
c	5	254	9.8	620	1	EXTN_TOBAC	P13983 nicotiana t
	6	248	9.8	463	1	YA68_MYCTU	O53416 mycobacteri
	7	248	9.8	778	1	YQ34_MYCTU	P71933 mycobacteri
	8	241.5	9.5	957	1	Y278_MYCTU	P56877 mycobacteri
	9	240	9.5	1901	1	YZ08_MYCTU	O53553 mycobacteri
	10	238.5	9.4	515	1	YI40_MYCTU	Q50594 mycobacteri
c	11	236	9.1	426	1	EXLP_TOBAC	Q03211 nicotiana t
	12	235.5	9.3	914	1	WA22_MYCTU	O06794 mycobacteri
	13	235	9.3	543	1	YP91_MYCTU	Q50630 mycobacteri
	14	233.5	9.2	338	1	GRP_ARATH	P27483 arabidopsis
	15	233	9.2	603	1	YD25_MYCTU	Q10637 mycobacteri
	16	232	9.2	801	1	Y747_MYCTU	O53810 mycobacteri
	17	229	9.0	491	1	YK98_MYCTU	Q10707 mycobacteri
	18	214.5	8.5	384	1	GRP1_PETHY	P09789 petunia hyb
	19	213.5	8.4	481	1	LORI_MOUSE	P18165 mus musculu
	20	213.5	8.4	747	1	SPD1_NEPCL	P19837 nephila cla
c	21	212	8.2	1567	1	FMN2_MOUSE	Q9j104 mus musculu
	22	210	8.3	1156	1	GLH4_CAEEL	O76743 caenorhabdi
c	23	207.5	8.0	817	1	VRP1_YEAST	P37370 saccharomyc
c	24	204	7.9	1255	1	DIA1_MOUSE	O08808 mus musculu
	25	203.5	8.0	465	1	GRP2_PHAVU	P10496 phaseolus v
	26	203	8.0	641	1	EBN1_EBV	P03211 epstein-bar
c	27	202.5	7.8	485	1	SSGP_VOLCA	P21997 volvox cart
	28	202.5	8.0	498	1	YI18_MYCTU	Q50615 mycobacteri
c	29	201	7.8	1248	1	DIA1_HUMAN	O60610 homo sapien
c	30	200.5	7.7	1509	1	GSR1_HUMAN	Q9nzm4 homo sapien
	31	199	7.9	937	1	HYR1_CANAL	P46591 candida alb
	32	197.5	7.8	344	1	POPA_RALSO	Q9rbs0 ralstonia s
c	33	195.5	7.5	267	1	EXTN_MAIZE	P14918 zea mays (m
c	34	195	7.5	449	1	APG_BRANA	P40603 brassica na
	35	193	7.6	183	1	GRP2_ORYSA	P29834 oryza sativ
c	36	193	7.4	865	1	CPN_DROME	Q02910 drosophila
	37	191	7.5	252	1	GRP1_PHAVU	P10495 phaseolus v
	38	191	7.5	316	1	LORI_HUMAN	P23490 homo sapien
	39	189	7.5	1113	1	N116_YEAST	Q02630 saccharomyc
	40	186	7.3	1387	1	TROP_HUMAN	Q12816 homo sapien
c	41	186	7.2	1955	1	PC15_HUMAN	Q96qu1 homo sapien
	42	184.5	7.3	419	1	CSP_PLACM	P08676 plasmodium
c	43	184.5	7.1	534	1	APG_ARATH	P40602 arabidopsis
c	44	183	7.1	743	1	EXT2_ARATH	Q9m1g9 arabidopsis
c	45	182.5	7.0	306	1	EXTN_DAUCA	P06599 daucus caro

ALIGNMENTS

```
RESULT 1
HRPN_ERWAM
ID   HRPN_ERWAM     STANDARD;      PRT;   403 AA.
AC   Q01099;
DT   01-APR-1993 (Rel. 25, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Harpin (Harpin-EA).
GN   HRPN.
OS   Erwinia amylovora.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Erwinia.
OX   NCBI_TaxID=552;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
RC   STRAIN=Ea321;
RX   MEDLINE=92320301; PubMed=1621099;
RA   Wei Z.-M., Laby R.J., Zumoff C.H., Bauer D.W., He S.Y., Collmer A.,
RA   Beer S.V.;
RT   "Harpin, elicitor of the hypersensitive response produced by the
RT   plant pathogen Erwinia amylovora.";
RL   Science 257:85-88(1992).
RN   [2]
RP   REVISIONS.
RC   STRAIN=Ea321;
RA   Laby R.J., Kim J.F., Beer S.V.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT
CC       UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO
CC       REQUIRED FOR PATHOGENICITY IN HOST PLANTS.
CC   -!- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M92994; AAC31644.2; -.
KW   Hypersensitive response.
```

```
FT   DOMAIN        1    270       GLY-RICH.
FT   DOMAIN       63     70       POLY-MET.
SQ   SEQUENCE   403 AA;  39697 MW;  146FA642351D8E87 CRC64;

Alignment Scores:
Pred. No.:              1.14e-121       Length:         403
Score:                  2050.00         Matches:        399
Percent Similarity:     99.01%          Conservative:   0
Best Local Similarity:  99.01%          Mismatches:     4
Query Match:            80.96%          Indels:         0
DB:                     1               Gaps:           0

SEQ12-SEQ4 (1-1390) x HRPN_ERWAM (1-403)

Qy      175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy      235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy      295 GCACTGGGGCTGGCGGCCGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy      355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy      415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGCACTGTCGAAC 474
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluLeuSerAsn 100

Qy      475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACAGGCTGGGCTCGAAAGCGGCAACAAT 534
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      101 AlaLeuAsnAspMetLeuGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy      535 ACCACTTCAACAACAAATTCCCCGCTGCACCAGCGCCTGGGTATTAACTCAACGTCCCAA 594
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy      595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGCACTCCAGCGACCCGATGCAGCAG 654
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy      655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy      715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy      775 GGAGTCACTGATCGCCTGTCGGCCTGATGGGTAATGCTCTGACCCAGCTCCTTGGCAAC 834
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuGlyAsn 220

Qy      835 GGGGGACTGGGAGGTGGTCAGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy      895 GGCGGCAAAGGCGTGCAAAACCTGAGCGGCGCCGGTGGCATCACCAGCAGTTAGGTAACGCC 954
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy      955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy     1015 AGGCACAGTTCAACCCGTTCTTCGTCAATCAAAGGCGATGGGCGAAGGAAATC 1068
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      281 SerAspSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300
```

```
Qy     1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGCCCG 1134
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy     1135 GGTCAGCAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGACCAAGCCAGATGAC 1194
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy     1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysSer 360

Qy     1255 CCCATGGCGGGTGATACCGGCCAACCGCAGGCACGCGGTCCGGTGGTTCTTCG 1314
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 AlaMetAlaGlyAspThrGlyAsnGlyAsnArgGlyAlaGlyGlyLeuSerSer 380

Qy     1315 CTGGGTATTGATGCCATGATGGCCGTGATCCCATTAACAATATGGCACTTGGCAAGCTG 1374
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy     1375 GGCGCCGCT 1383
            |||||||||
Db      401 GlyAlaAla 403

RESULT 2
HRPN_ERWCA
ID   HRPN_ERWCA      STANDARD;      PRT;   356 AA.
AC   Q47279;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Harpin (Harpin-ECC) (Fragment).
GN   HRPN.
OS   Erwinia carotovora.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pectobacterium.
OX   NCBI_TaxID=554;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=71;
RX   MEDLINE=96405946; PubMed=8811071;
RA   Cui Y., Madi L., Mukherjee A., Dumenyo C.K., Chatterjee A.K.;
RT   "The RsmA-mutants of Erwinia carotovora subsp. carotovora strain
RT   Ecc71 overexpress hrpNEcc and elicit a hypersensitive reaction-like
RT   response in tobacco leaves.";
RL   Mol. Plant Microbe Interact. 9:565-573(1996).
CC   -!- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT
CC       UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO
CC       REQUIRED FOR PATHOGENICITY IN HOST PLANTS.
CC   -!- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L78834; AAB49733.1; -.
KW   Hypersensitive response.
FT   DOMAIN        1     224       GLY-RICH.
FT   NON_TER     356     356
SQ   SEQUENCE   356 AA;  35621 MW;  108B468B9D27F9DE4 CRC64;

Alignment Scores:
Pred. No.:              5.61e-45        Length:         356
Score:                  828.50          Matches:        196
Percent Similarity:     57.25%          Conservative:   37
Best Local Similarity:  48.16%          Mismatches:     115
Query Match:            32.72%          Indels:         59
DB:                     1               Gaps:           11
```

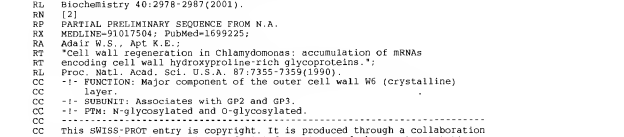
```
SEQ12-SEQ4 (1-1390) x HRPN_ERWCA (1-356)

Qy    181 CTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGA 240
          ||||||:::   |||   |||||||||   :::||||||:::|||       |||||||||
Db      2 LeuAsnSerLeuGlyGlyGlyAlaSer---LeuGlnIleThrIle---LysAlaGlyGly 19

Qy    241 AATAACGGGTTGCTG---GGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTGCA 297
          |||   ||||||            |||   ||||||   |||          ||||||
Db     20 AsnGlyGlyLeuPheProSerGlnSerSerGlnAsnGlyGlySerProSerGlnSerAla 39

Qy    298 CTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACC 357
                      ||||||   :::::::::         |||||||:::   ::::::|||
Db     40 Phe---------GlyGlyGlnArgSerAsnIleAlaGluGlnLeuSerAspIleMetThr 56

Qy    358 GGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGC 417
             ||||||   |||   |||||||||||||||||         |||   |||||||||||
Db     57 ThrMetMetPheMetGlySerMetMetGlyGlyGly------MetSerGlyGlyLeuGly 74

Qy    418 -----------------GGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGC 459
                           |||||||||||   ||||||   |||   |||||||||||
Db     75 GlyLeuGlySerSerLeuGlyGlyLeuGlyGlyGlyLeuLeuGly---GlyGlyLeuGly 93

Qy    460 GAAGGACTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCG 519
             ||||||   :::::::|||         ||||||   :::|||         |||
Db     94 GlyGlyLeuGlySerSerLeuGlySerGlyLeuGlySerAlaLeuGly------GlyGly 111

Qy    520 AAAGGCGGCAACAATACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATT 579
             ||||||                                    |||||||||
Db    112 LeuGlyGly---------------------------------------AlaLeuGlyAla 118

Qy    580 AACTCAACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGC 639
                    :::   |||      :::      |||:::
Db    119 GlyMetAsnAlaMetAsnProSerAlaMetMetGlySer--------------------- 131

Qy    640 GACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGAT 699
                                 :::         ::::::   |||   |||
Db    132 -----------------------LeuLeuPheSerAlaLeuGluAspLeuLeuGlyGly 143

Qy    700 GGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAG 759
          |||         ||||||         |||   ||||||||||:::         |||
Db    144 GlyMetSerGlnGlnGlnGlyGlyLeuPheGlyAsnLysGlnProSerSerProGluIle 163

Qy    760 AACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGC 819
          :::||||||   :::||||||   |||||||||||   ::::::|||||||||||||||
Db    164 SerAlaTyrThrGlnGlyValAsnAspAlaLeuSerAlaIleLeuGlyAsnGlyLeuSer 183

Qy    820 CAGCTCCTTGGCAACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTT 879
          |||      |||
Db    184 GlnThrLysGlyGln------------------------------------------Thr 189

Qy    880 GACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAG 939
                      ||||||   |||||||||   |||||||||         :::
Db    190 SerProLeuGlnLeuGlyAsnAsnGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsn 209

Qy    940 CAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAAT 999
          |||||||||:::   :::|||      :::|||   ||||||||||:::|||   ||||||
Db    210 GlnLeuGlySerThrLeuGlyMetSerValGlyGlnLysAlaGlyLeuGlnGluLeuAsn 229

Qy   1000 GATATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCG 1059
          :::|||   ||||||         |||   ||||||   ||||||:::|||   ||||||
Db    230 AsnIleSerThrHisAsnAspSerProThrArgTyrPheValAspLysGluAspArgGly 249

Qy   1060 ATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAG 1119
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||   :::
Db    250 MetAlaLysGluIleGlyGlnPheMetAspGlnTyrProGluValPheGlyLysAlaGlu 269

Qy   1120 TACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTG 1179
          ||||||||         |||      |||   :::||||||||||||||||||||||||
Db    270 TyrGlnLysAspAsnTrpGlnThrAlaLysGlnGluAspLysSerTrpAlaLysAlaLeu 289

Qy   1180 AGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAG 1239
          |||||||||||||||||||||||||         ||||||:::::::|||   ||||||
Db    290 SerLysProAspAspAspGlyMetThrLysGlySerMetAspLysPheMetLysAlaVal 309

Qy   1240 GGCATGATCAAAAGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGT 1299
          ||||||||||||         :::   |||||||||||||||   ||||||   |||||||||
Db    310 GlyMetIleLysSerAlaIleArgGlyAspThrGlyAsnThrAsnLeuSerAlaArgGly 329

Qy   1300 GCCGGTGGTTCTTCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATG 1359
             ||||||:::||||||||||||||||||   |||   ||||||   |||   ||||||
Db    330 AsnGlyGlyAlaSerLeuGlyIleAspAlaAlaMetIleGlyAspArgIleValAsnMet 349

Qy   1360 GCACTTGGCAAGCTGGGCGCG 1380
             |||   ||||||   :::
Db    350 GlyLeuLysLysLeuSerSer 356
```

```
RESULT 3
HRPN_ERWCH
ID   HRPN_ERWCH     STANDARD;      PRT;   340 AA.
AC   Q47278;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Harpin (Harpin-ECH).
GN   HRPN.
OS   Erwinia chrysanthemi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pectobacterium.
OX   NCBI_TaxID=556;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EC16 / AC1450;
RX   MEDLINE=96172740; PubMed=8589405;
RA   Bauer D.W., Wei Z.M., Beer S.V., Collmer A.;
RT   "Erwinia chrysanthemi harpinEch: an elicitor of the hypersensitive
RT   response that contributes to soft-rot pathogenesis.";
RL   Mol. Plant Microbe Interact. 8:484-491(1995).
CC   -!- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT
CC       UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO
CC       REQUIRED FOR PATHOGENICITY IN HOST PLANTS.
CC   -!- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L39897; AAC31978.1; -.
KW   Hypersensitive response.
FT   DOMAIN        1    203       GLY-RICH.
SQ   SEQUENCE   340 AA;  34274 MW;  E4D82C23731EF4C5 CRC64;

Alignment Scores:
Pred. No.:              4.52e-38      Length:         340
Score:                  718.50        Matches:        172
Percent Similarity:     52.72%        Conservative:   41
Best Local Similarity:  42.57%        Mismatches:     114
Query Match:            28.38%        Indels:         77
DB:                     1             Gaps:           10

SEQ12-SEQ4 (1-1390) x HRPN_ERWCH (1-340)

Qy    211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
          |||||||||:::|||         ||||||:::   |||:::   |||
Db      1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyLeu 20

Qy    268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
                   ||||||   |||   |||||||||         ||||||
Db     21 GlyAlaGlnGlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 40

Qy    316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                :::   |||:::::::::|||         ||||||   ||||||
```

RESULT 4
GP1_CHLRE
ID GP1_CHLRE STANDARD; PRT; 555 AA.
AC Q9FPQ6; Q03927;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gp1 precursor (Hydroxyproline-rich
DE glycoprotein 1).
GN GP1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21159092; PubMed=11258910;
RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA Goodenough U.H.;
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
RT plant hydroxyproline-rich glycoproteins.";
RL Biochemistry 40:2978-2987(2001).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.E.;
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
RT encoding cell wall hydroxyproline-rich glycoproteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC -!- FUNCTION: Major component of the outer cell wall W6 (crystalline)
CC layer.
CC -!- SUBUNIT: Associates with GP2 and GP3.
CC -!- PTM: N-glycosylated and O-glycosylated.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AF309494; AAG45420.1; -.
DR EMBL; M58496; AAA69706.1; ALT_SEQ.
DR GlycoSuiteDB; Q9FPQ6; -.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR003882; Pistil_extensin.
DR PRINTS; PR01217; PRICHEXTENSN.
DR PRINTS; PR01218; PSTLEXTENSIN.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GP1.
FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSPX REPEATS.
FT DOMAIN 259 279 POLY-PRO.
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465302PS CRC64;

Alignment Scores:
Pred. No.: 7.94e-13 Length: 555
Score: 316.50 Matches: 129
Percent Similarity: 34.11% Conservative: 18
Best Local Similarity: 29.93% Mismatches: 142
Query Match: 12.21% Indels: 142
DB: 1 Gaps: 18

SEQ12-SEQ4 (1-1390) x GP1_CHLRE (1-555)



```
Qy   1322 ATA------------------------------------CCCAGCGAAGAACCACCG 1302
                                                   ||||||     ||||||
Db     27 AlaAsnAlaGlnCysValProGlyGlyIlePheAsnCysProProSerProAlaProPro 46

Qy   1301 GCACCGCGT-----------GCCTGCAGGTTGCCGTTGCCGGTATCACCCGCCATGGGC 1254
           :::|||              |||          |||   |||   |||||||||
Db     47 SerProAlaProProSerProAlaProProSerProAlaProProSerProAla------ 64

Qy   1253 CTTTTGATCATGCCCTTGGCTTTGTTGAACTGCTCCATACTGGCTGGTGTCATTCCGTCG 1194
                                                                |||
Db     65 ------------------------------------------------------ProPro 66

Qy   1193 TCATCTGGCTTGCTCAGTGCTTTTGCCCATGATTTGTCATCGGTTTTCACCTCCTGACCC 1134
           |||  |||      |||      |||                |||||||||   ||
Db     67 SerProGlyProProSerPro-AlaProPro--------------SerProProSerPr 81

Qy   1133 GGGCCTTTCTGGTACTGCGGCTTGCCAAACACCTCAGGATACTGGTCCATGAACTGACCG 1074
           |                |||           |||           |||   :::
Db     81 oAlaProProSerProAla-----------ProProSerProAlaProProSerProAl 97

Qy   1073 ATTTCCTTCGCCATCGCCCGATCGCCTTTATTGACGAAAGAACGGGTTGAACTGTGCCTG 1014
             |||||||||:::|||
Db     97 aProProSerProAlaProPro------------------------------------- 104

Qy   1013 TGCGTACCGATATCATTCAGCGCCTGAATGCCCGCTTTCATACCGATACCGGTACCCACG 954
                         ||||||      |||         |||    |||    |||
Db    105 ------------SerProAlaProProSerProAlaPro--ProSerProAlaProPro 119

Qy    953 GCGTTACCTAACTGCTGGTAGTCCACCGGCCCGCTCAGGTTTTGCAGCCCTTTGCCGCCC 894
           :::   |||             |||        |||                |||   ||||||
Db    120 SerProPro-----------SerProAlaProProSerProSerProProAlaProPro 135

Qy    893 AGCGACGAACCGTCAAGACCCGTGCCAGCATTACCGCCC---------TGACCACCTCCC 843
           |||       |||       |||    |||    |||||||||         |||    |||
Db    136 SerProSerProProSerProAlaProProLeuProProSerProAlaProProSerPro 155

Qy    842 AGTCCCCCGTTGCCAAGGAGCTGGCTCAGACCATTACCCATCAGGCCCGACAGCGCATCA 783
           |||||||||:::|||     |||             |||:::|||        |||           |||
Db    156 SerProProValProProSerProSerProProValProProSerProAlaProProSer 175

Qy    782 GTGACTCCTTTTTTATAGGCGTTCTGCTCGCCTTCG-----------GTCGGCTGCTTG 735
             ||||||                   |||||||||
Db    176 ProThrPro--------------ProSerProSerProProValProProSerProAla 190

Qy    734 CCCCCAGAGGAACTGCCCTGGGTGCCATCTTGCCCATCACCAAACAGGCTTTGCATTATC 675
           ||||||           |||    ||||||       |||:::|||
Db    191 ProProSerProAlaProProValProProSerProAlaPro----------------- 204

Qy    674 TCGCTGAACATCTTCAGCAGCTGCTGCATCGGGTCGCTGGAGTCTGAGGTGGAATCTGTG 615
                                                   |||              |||
Db    205 ---------------------------------------ProSerProAlaProProVal 211

Qy    614 CCGGAGGTGGAATCGTCGTTTTGGGACGTTGAGTTAATACCCAGCGCCTGGTCCAGCGGG 555
           |||         :::                            ||||||       |||
Db    212 ProProSerProAlaPro--------------------ProSerProProSerProAla 224

Qy    554 GAATTTGTTGTTGAAGTGGTATTGTTGCCGCCTTTCGAGCCCAGCGTGTTCAGCGAACCG 495
                                           ||||||      ||||||              |||
Db    225 --------------------------ProProSerProProSer------ProAlaPro 233

Qy    494 CCTAACATATCGTTCAGCGCGTTCGACAGT-----------------------CCTTCG 459
           |||:::    |||       |||       |||                        ||||||
Db    234 ProSerProSerProProAlaProProSerProValProProSerProAlaProProSer 253

Qy    458 CCCAGGCCACCTGAGCCACCCAAGCCATTACCTAAGCCACCGCCTAAGCCACCGCCCATC 399
           |||   ||||||    |||        |||    |||    |||||||||    |||||||||
Db    254 ProAlaProProSerProLysProProAlaProProProProSerProProProPro 273

Qy    398 AGCCCACCACCGCCCATCATGCTCATCATCATCATCATGCCGGTGAGTAAGCCAGCCAGC 339
              ||||||     |||                     ||||||    |||    |||:::
Db    274 ProProProArgProProPheProAlaAsnThrProMetProProSerProProSer--- 292

Qy    338 TGATTGACGGTATCATTTTGATTACCGCCGCCCAGCCCCAGTGCAGAATTGCCACCCAAC 279
                                         |||||||||||||||:::            ||||||
Db    293 ------------------------ProProProSerProAlaProProThrProProThr 304

Qy    278 CCAGCATTCTGGCGACTGGTACCCAGCAACCCGTTATTTCCGCCCGCACCGCCGATAGAA 219
              |||                     ||||||     |||              |||     ||||||
Db    305 ProPro------------SerProSerProPro------SerProValProProSerPro 318

Qy    218 ATTTGCATCGTTGACGCTCCCAGCCCA 192
                :::          :::|||:::|||
Db    319 AlaProValProProSerProAlaPro 327
```

```
RESULT 5
EXTN_TOBAC
ID   EXTN_TOBAC     STANDARD;      PRT;   620 AA.
AC   P13983;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
GN   HRGPNT3.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Xanthi; TISSUE=Leaf;
RX   MEDLINE=90128263; PubMed=2612909;
RA   Keller B., Lamb C.J.;
RT   "Specific expression of a novel cell wall hydroxyproline-rich
RT   glycoprotein gene in lateral root initiation.";
RL   Genes Dev. 3:1639-1646(1989).
CC   -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC       THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC       MAIN ROOT.
CC   -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC   -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC       SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC       GLYCOSYLATED.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X13885; CAA32090.1; -.
DR   PIR; S06733; S06733.
KW   Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW   Hydroxylation.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    620       EXTENSIN.
FT   REPEAT       70     73       H-A-P-P.
FT   REPEAT      148    151       H-A-P-P.
FT   DOMAIN      229    242       2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
FT   REPEAT      229    235       1.
FT   REPEAT      236    242       2.
FT   DOMAIN      205    620       CONTAINS THE SER-PRO(4) REPEATS.
FT   DOMAIN      499    600       3 X APPROXIMATE TANDEM REPEATS.
SQ   SEQUENCE   620 AA;  65406 MW;  641DD2278AB28524 CRC64;

 Alignment Scores:
 Pred. No.:               6.69e-09      Length:        620
 Score:                   254.00        Matches:       113
 Percent Similarity:      34.13%        Conservative:  29
 Best Local Similarity:   27.16%        Mismatches:    150
 Query Match:             9.80%         Indels:        124
 DB:                      1             Gaps:          20
```

```
SEQ12-SEQ4 (1-1390) x EXTN_TOBAC (1-620)

Qy   1339 CGGCCATCATGGCATCAATACCCAGCGAAGAACCACCGGCACCGCGTGCCTGCAGGTTGC 1280
             :::|||      |||:::::          ||||||||        |||
Db    226 GlnProProThrHisArgHisAlaProProThrHisArgHisAlaProProThr------ 243

Qy   1279 CGTTGCCGGTATCACCCGCCATGGGCCTTTGATCATGCCCTTGGCTTTGTTGAACTGCT 1220
                      |||   |||
Db    244 ------------HisGlnPro--------------------------------------- 246

Qy   1219 CCATACTGGCTGGTGTCATTCCGTCGTCATCTGGCTTGCTCAGTGCTTTTGCCCATGATT 1160
                            |||       ||||||||       |||
Db    247 ---------------SerProLeuArgHisLeuProProSerProArgArgGlnProGln 261

Qy   1159 TGTCATCGGTTTTCACCTCCTGACCCGGGCCTTTCTGGTACTGCGGCTTGCCAAACACCT 1100
                :::|||||||||   |||                  |||    |||:::|||
Db    262 ProProThrTyrSerProProProPro---------------AlaTyrAlaGlnSerPro 276

Qy   1099 CAGGATACT---------------GGTCCATGAACTGACCGATTTCCTTCGCCATCGCCC 1055
              |||   :::                |||   |||        |||   |||||||||
Db    277 GlnProSerProThrTyrSerProProProProThrTyrSerProProProProSerPro 296

Qy   1054 GATCGCCTTTATTGACGAAAGAACGGGTTGAACTGTGCCTGTGCGTACCGATATCATTCA 995
                                                                 :::|||
Db    297 ---------------------------------------------------IleTyrSer 299

Qy    994 GCGCCTGAATGCCCGCTTTCATACCGATACCGGTACCCACGGCGTTACCTAACTGCTGGT 935
              |||     ||||||:::   |||   |||    ||||||       |||
Db    300 ProProPro--ProAlaTyrSerProSerProProProThrProThrProThrPhe---- 317

Qy    934 AGTCCACCGGCCCGCTCAGGTTTTGC---------------AGCCCTTTGCCGCCCAGCG 890
               |||     |||       :::                     ||||||    ||||||:::
Db    318 --SerProProProProAlaTyrSerProProProThrTyrSerProProProProThrT 337

Qy    889 ACGAACCG--------------------TCAAGACCCGTGCCAGCATTACCGCCCTGAC 851
                |||                       |||   |||   |||        |||   |
Db    337 yrLeuProLeuProSerSerProIleTyrSerProProProProValTyrSerProProP 357

Qy    850 CACCTCCC------AGTCCCCCGTTGCCAAGGAGCTGGCTCAGACCATTACCCATCAGGC 797
              ||||||||      |||||||||   |||             |||   |||       |
Db    357 roProProSerTyrSerProProProProThrTyrLeuProProProProProSerSerP 377

Qy    796 CCGACAGCGCATCAGTGACTCCTTTTTTATAGGCGTTCTGCTCGCCTTCGGTCGGCTGCT 737
              ||          |||    :::|||                  |||:::
Db    377 roProProProSerPheSerPro---------------ProProProThrTyrGluGlnS 392

Qy    736 TGCCCCCAGAGGAACTGCCCTGGGTGCCATCTTGCCCATCACCAAACAGGCTTTGCATTA 677
              ||||||        :::    |||        |||:::|||
Db    392 erProProProProProAlaTyrSerProProLeuProAlaProProThrTyr------- 409

Qy    676 TCTCGCTGAACATCTTCAGCAGCTGCTGCATCGGGTCGCTGGAGTCTGAGGTGGAATCTG 617
                                                |||            |||
Db    410 ----------------------------------SerProProProProThrTyrSerP 418

Qy    616 TGCCGGAGGTGGAATCGTCGTTTTGGGACGTTGAGTTAATACCCAGCGCCTGGTCCAGCG 557
               |||             ::::::              :::|||       :::|||
Db    418 roPro------ProProThrTyrAlaGlnProProProLeuProProThrTyrSer---- 434

Qy    556 GGGAATTTGTTGTTGAAGTGGTATTGTTGCCGCCTTTCGAGCCCAGCGTGTTCAGCGAAC 497
                                            ||||||    |||      :::|||    |
Db    435 ----------------------------ProPro------ProProAlaTyrSerProP 443

Qy    496 CGCCTAACATATCGTTCAGC--------------------------------GCGTTCG 470
              |||||     ::::::|||                                  |||:::
Db    443 roProProProThrTyrSerProProProProThrTyrSerProProProProAlaTyrA 463

Qy    469 ACAGTCCTTCGCCCAGGCCACCT-----GAGCCACCCAAGCCATTACCTAAGCCACCGC 416
                 |||   |||   ||||||            ||||||   |||       |||||||
Db    463 laGlnProProProProProProThrTyrSerProProProProAlaTyrSerProProP 483

Qy    415 CTAAGCCACCGCCCATCAGCCCACCACCGCCCATCATGCTCATCATCATCATCATGCCGG 356
              ||      |||        ||||||||||||||          :::      :::    |||
Db    483 roProSerProIleTyrSerProProProPro------GlnValGlnProLeuProProT 501

Qy    355 TGAGTAAGCCAGCCAGCTGATTGACGGTATCATTTTGATTACCGCCGCCCAGCCCCAGTG 296
                   |||               :::           |||||||||||
Db    501 hrPheSerProProProProArgArgIleHis------LeuProProProProHisArgG 519

Qy    295 CAGAATTGCCACCCAACCCAGCATTCTGGCGACTGGTACCCAGCAACCCGTTATTTCCGC 236
                   ||||||    |||    :::     :::        ||||||    |||    |||    |
Db    519 lnProArgProProThrProThrTyrGlyGln---ProProSerProProThrPheSerP 538

Qy    235 CCGCACCGCCGATAGAAATTTGCATCGTTGACGCTCCCAGCCCA 192
               ||    ||||||    :::|||              :::|||    |||
Db    538 roProProProArgGlnIle---------HisSerProProPro 549

RESULT 6
YA68_MYCTU
ID   YA68_MYCTU     STANDARD;      PRT;   463 AA.
AC   O53416;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv1068c.
GN   RV1068C OR MT1097 OR MTV017.21C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL021897; CAA17184.1; -.
DR   EMBL; AE006991; AAK45353.1; ALT_INIT.
DR   TIGR; MT1097; -.
DR   TubercuList; Rv1068c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT    218    218       T -> S (IN REF. 2).
FT   CONFLICT    235    235       G -> GQGGAGIGGADGTKGGDAGAGGAGGAGGWIHGHGGV
```

```
FT                                GGDGGTGGQGGDGVQGEPGDTGAAGGAGG (IN REF.
FT                                2).
SQ   SEQUENCE   463 AA;  39305 MW;  CF5696A7E9593952 CRC64;

Alignment Scores:
Pred. No.:              1.58e-08        Length:         463
Score:                  248.00          Matches:        93
Percent Similarity:     37.67%          Conservative:   17
Best Local Similarity:  31.85%          Mismatches:     112
Query Match:            9.79%           Indels:         70
DB:                     1               Gaps:           13

SEQ12-SEQ4 (1-1390) x YA68_MYCTU (1-463)

Qy     196 CTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAATAACGGGTTGCTG 255
            :::||||||      :::                ||||||||||||||     ||||||
Db     203 IleGlyAlaProGlyValAlaGlyGlyAlaGlyGlyAlaGlyGlyThrAlaGlyLeuPhe 222

Qy     256 GGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTGCACTGGGGCTGGGCGGCGGT 315
           |||           |||   |||   ||||||           |||       ||||||
Db     223 Gly------------AsnGlyGlyAlaGlyGlyAlaGlyGlyAlaGlyGlyAlaGlyGly 238

Qy     316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                    |||                ||||||    |||:::|||
Db     239 ArgGlyGlyAsp------GlyGlySerAlaGlyTrpLeuSerGlyAsnGlyGlyAspAla 256

Qy     376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                    |||||||||||      |||    |||      ||||||
Db     257 GlyThrGlyGlyGlyGlyGlyAsnAlaGlyAsnGlyGlyAsnGlyGlySerAlaGlyTrp 276

Qy     436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
           |||   |||:::||||||   |||   |||                            |||
Db     277 LeuSerGlyAsnGlyGlyThrGlyGlyGly---------------------------Gly 287

Qy     496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
           |||:::             |||   :::|||||||||                  ||||||
Db     288 GlyThrAlaGlyAla---GlyGlyGlnGlyGlyAsnGly---------------AsnSer 301

Qy     556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
               :::|||             |||                |||  .     :::|||
Db     302 GlyIleAspProGly---------AsnGlyGlyGlnGlyAlaAspThrGlyAsnAlaGly 318

Qy     616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
                             |||:::
Db     319 AsnGlyGlyHisGlyGlySerAla------------------------------------ 326

Qy     676 ATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACCCAGGGCAGTTCC--------- 726
                       |||||||||||||||     |||      |||
Db     327 ------AlaLysLeuPheGlyAspGlyGlyAlaGlyGlyAlaGlyGlyMetGlySerThr 344

Qy     727 --------------------------TCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAG 759
                                     ||||||                 |||
Db     345 GlyGlyThrGlyGlyGlyGlyGlyPheGlyGlyGlyThrGlyGlyAsnGlyGlyAsnGly 364

Qy     760 AACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGC 819
           :::|||          |||   :::           :::||||||:::|||:::|||   |||
Db     365 HisAlaGlyGlyAlaGlyGlySerGlyGlyThrAlaGlyLeuLeuGlySerGlyGlySer 384

Qy     820 CAGCTCCTT------------GGCAACGGGGGACTGGGAGGTGGTCAGGGC------GGT 861
                                 |||||||||||||||||   |||   |||         |||
Db     385 GlyGlyThrGlyGlyAspGlyGlyAsnGlyGlyLeuGlyAlaGlySerGlyAlaLysGly 404

Qy     862 AATGCTGGCACGGGTCTTGACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGC 921
           |||   |||   |||   |||               ||||||||||||
Db     405 AsnGlyGlyAsnGlyGlyAsp-----------GlyGlyLysGly--------------- 415

Qy     922 GGGCCGGTGGACTACCAGCAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCG 981
                    |||   |||   :::||||||             |||   |||   |||
Db     416 ------GlyAspAlaGlnLeuIleGlyAsnGlyGlyAsnGlyGlyAsnGlyGlyLysGly 433

Qy     982 GGC---------ATTCAGGCGCTGAATGATATCGGT 1008
           |||          :::      :::|||      |||
Db     434 GlyThrGlyLeuMetProGlyIleAsnGlyThrGly 445
```

```
RESULT 7
YQ34_MYCTU
ID   YQ34_MYCTU     STANDARD;      PRT;   778 AA.
AC   P71933;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv2634c.
GN   RV2634C OR MT2712 OR MTCY441.04C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z80225; CAB02341.1; -.
DR   EMBL; AE007103; AAK47026.1; ALT_INIT.
DR   TIGR; MT2712; -.
DR   TubercuList; Rv2634c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT     51     51       V -> L (IN REF. 2).
FT   CONFLICT     63     63       Q -> H (IN REF. 2).
FT   CONFLICT    274    274       A -> T (IN REF. 2).
SQ   SEQUENCE   778 AA;  63131 MW;  DAB20FE58E4999E7 CRC64;

Alignment Scores:
Pred. No.:              1.6e-08         Length:         778
Score:                  248.00          Matches:        126
Percent Similarity:     31.93%          Conservative:   26
Best Local Similarity:  26.47%          Mismatches:     169
Query Match:            9.79%           Indels:         156
DB:                     1               Gaps:           19

SEQ12-SEQ4 (1-1390) x YQ34_MYCTU (1-778)
```

```
Qy   193 GGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGC------GGAAATAAC 246
Db   353 GlyAsnGlyGlyAsnAlaAsnTrpPheGlySerGlyGlyAlaGlyGlyGlnGlyGlyThr 372

Qy   247 GGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTCCACTCGGGCTG 306
Db   373 GlyLeuAlaGlyThrAsnGlyValAsnProGlySerIleAlaAsnProAsnThrGlyAla 392

Qy   307 GGCGGC------------GGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
Db   393 AsnGlyThrAspAsnSerGlyAsnGlyAsnGlnThr---------------------- 404

Qy   355 ACCGGCATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
Db   405 ----------------------------GlyGlyAsnGly------GlyProGlyPro 412

Qy   415 GGCGGTGGCTTAGGT------AATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTG 468
Db   413 AlaGlyGlyValGlyGluAlaGlyGlyValGlyGlyGlyLeuGlyGlySerLeuAsp 432

Qy   469 TCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGCTCGAAAGCGGC 528
Db   433 AspGly------AsnAspGlyThrGlyGly---------------------LysGlyGly 443

Qy   529 AACAATACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACG 588
Db   444 AlaGlyGlyThrAlaGlyThrAspGlyGlyAlaGlyGlyAlaGlyGlyAlaGlyGlyIle 463

Qy   589 TCCCAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATG 648
Db   464 GlyGluThrAspGlySerAlaGlyGlyVal---------------------------- 473

Qy   649 CAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGAT 708
Db   474 ----------------------------------------------AlaThrGlyGlyGlyGlyAsp 481

Qy   709 GGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGGCAG---CAGAACGCC 765
Db   482 GlyAlaThrGlyGlyValAspGlyGlyValGlyGlyAlaGlyGlyLysGlyGlyGlnGly 501

Qy   766 TATAAAAAAGGAGTCACTGATGCGCTGTCGGGC---------------------------- 798
Db   502 HisAsnThrGlyValGlyAspAlaPheGlyGlyAspGlyGlyIleGlyGlyAspGlyAsn 521

Qy   799 ----------------CTGATGGGTAATGGTCTGAGC---------------------- 819
Db   522 GlyAlaLeuGlyAlaAlaGlyGlyAsnGlyGlyThrGlyGlyAlaGlyAsnGlyGly 541

Qy   820 ---------CAGCTCCTTGGCAACGGGGGACTGGGAGGTGGTCAGGCGGTAATGCTGGC 870
Db   542 ArgGlyGlyMetLeuIleGlyAsnGlyGly---AlaGlyGlyAlaGlyGlyThrGlyGly 560

Qy   871 ACGGGTCTTGACGGTTCGTCG---------------------CTGGGCGGCAAAGGG 906
Db   561 ThrGlyGlyGlyGlyAlaAlaGlyPheAlaGlyGlyValGlyGlyAlaGlyGlyGluGly 580

Qy   907 CTGCAAAACCTGAGCGGGCCGGTGAC----------TACCAGCAGTTAGGTAAC 951
Db   581 LeuThrAspGlyAlaGlyThrAlaGluGlyGlyThrGlyGlyLeuGlyGlyLeuGlyGly 600

Qy   952 GCCGTGGGTACC---------------------------------------------- 963
Db   601 ValGlyGlyThrGlyGlyMetGlyGlySerGlyGlyValGlyGlyAsnGlyGlyAlaAla 620

Qy   964 ----------------------------------------GGTATCGGTATGAAAGCGGGC 984
Db   621 GlySerLeuIleGlyLeuGlyGlyGlyGlyAlaGlyGlyValGlyGlyThrGlyGly 640

Qy   985 ATTCAGGCGCTGAATGATATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAAT 1044
Db   641 IleGlyGlyIleGlyGlyAlaGlyGlyAsnGlyGlyAlaGlyGlyAlaGlyThrThrThr 660

Qy  1045 AAAGGCGATCG------GGCGATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCC 1097
Db   661 GlyGlyGlyAlaThrIleGlyGlyGlyGlyGlyThrGlyGlyValGlyAlaGlyAlaGly 680

Qy  1098 TGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGA 1157
Db   681 ThrGlyGlyThrGlyGlyAlaGlyGlyThrThrGlyGlySerGlyGlyAlaGly-GlyLe 700

Qy  1158 CAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTAT 1217
Db   700 uIleGlyTrpAlaGlyAlaAla----------------GlyGlyThrGlyAla----- 712

Qy  1218 GGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGGCCCATGGCGGGTGATACCGGCAA 1277
Db   713 ----------------------------------------GlyGlyThrGlyGlyGl 718

Qy  1278 CGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCGCTGGGTATTGATGCCATGATGGC 1337
Db   718 nGlyGlyLeuGlyGlyGlnGlyGlyAsnGlyGlyAsnGlyGlyThrGlyAla------Th 736

Qy  1338 CGGTGATGCCATTAACAATATGGCACTTGGCAAGCTGGGCGCGGCT 1383
Db   736 rGlyGlyGlnGlyGlyAspPheAlaLeuGlyGlyAsnGlyGlyAla 751
```

```
RESULT 8
Y278_MYCTU
ID   Y278_MYCTU      STANDARD;      PRT;   957 AA.
AC   P56877;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv0278c precursor.
GN   RV0278C OR MT0291 OR MTV035.06C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   -----------------------------------------------------------------------
DR   EMBL; AL021930; CAA17353.1; -.
DR   EMBL; AE006936; AAK44511.1; ALT_INIT.
DR   TIGR; MT0291; -.
DR   TubercuList; Rv0278c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    957       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                                RV0278C.
FT   CONFLICT     40     40       M -> I (IN REF. 2).
FT   CONFLICT    158    163       MISSING (IN REF. 2).
FT   CONFLICT    807    807       R -> G (IN REF. 2).
SO   SEQUENCE   957 AA;  81905 MW;  71EBABD417FBA47C CRC64;

Alignment Scores:
Pred. No.:              4.11e-08      Length:          957
Score:                  241.50        Matches:         117
Percent Similarity:     35.10%        Conservative:    42
Best Local Similarity:  25.83%        Mismatches:      149
Query Match:            9.54%         Indels:          145
DB:                     1             Gaps:            21

SEQ12-SEQ4 (1-1390) x Y278_MYCTU (1-957)

Qy    178 AGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGC 237
            :::|||          |||:::|||    :::                   |||||||||||
Db    362 ThrLeuThrAlaGlyGlyIleGlyGlyAla-----------------GlyGlyAlaGly 375

Qy    238 GGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGC--------- 288
            ||||||  |||||||||    |||       :::   ||||||
Db    376 GlyAsnAlaGlyLeuLeuPheGlySerGlyGlySerGlyGlyAlaGlyGlyPheGlyPhe 395

Qy    289 ---AATTCTGCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCT 345
               :::         |||     |||||||||                        |||
Db    396 AlaAspGlyGlyGlnGlyGlyProGlyGlyAsn-----------------------Ala 407

Qy    346 GGCTTACTCACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGC 405
            |||                            :::::  |||   ||||||      |||
Db    408 Gly--------------------------ThrValPheGlySerGlyGlyAlaGlyGly 418

Qy    406 ---GGTGGCTTAGGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTG------ 456
               ||||||:::|||   |||       |||:::||||||:::||||||
Db    419 AsnGlyGlyValGlyGlnGlyPheAlaGlyGlyIleGlyGlyAlaGlyGlyThrProGly 438

Qy    457 -----------------------GGCGAAGGACTGTCGAACGCGCTG---------AAC 483
                                   |||   |||   ::::::|||:::          |||
Db    439 LeuIleGlyAsnGlyGlyAsnGlyGlyAsnGlyGlyAlaSerAlaValThrGlyGlyAsn 458

Qy    484 GATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCA 543
                  :::||||||:::      :::|||:::   |||||||||       :::
Db    459 GlyGlyIleGlyGlyThrGlyValLeuIleGlyAsn---GlyGlyAsnGlyGlySerGly 477

Qy    544 ACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGAT 603
                 :::                   |||:::::                      |||
Db    478 GlyIleGlyAlaGlyLysAlaGlyValGlyGlyValSerGlyLeuLeuLeuGlyLeuAsp 497

Qy    604 TCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAG 663
                 :::         |||||||||                    |||   :::|||
Db    498 GlyPheAsnAlaProAlaSerThrSerProLeuHisThrLeuGlnGlnAsnValLeuAsn 517

Qy    664 ATGTTCAGCGAGATAATGCAAAGC---------------CTGTTTGGTGATGGGCAAGAT 708
            :::   :::|||     |||:::                 |||   |||:::|||    :::
Db    518 ValValAsnGluProPheGlnThrLeuThrGlyArgProLeuIleGlyAsnGlyAlaAsn 537

Qy    709 GGCACCCAGGGCAGT------------TCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAG 756
            ||||||   |||:::               :::||||||               |||
Db    538 GlyThrProGlyThrGlyAlaAspGlyGlyAlaGlyGlyTrpLeuPheGlyAsnGlyAla 557

Qy    757 CAGAACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGC---CTGATGGGTAATGGT 813
                               |||                |||   |||   |||||||||
Db    558 AsnGlyThrProGlyThrGlyAlaAlaGlyGlyAlaGlyGlyTrpLeuPheGlyAsnGly 577

Qy    814 CTGAGCCAGCTCCTTGGCAACGGGGGACTGGGAGGT----------------------- 849
                                  ||||||||||||    |||
Db    578 --------------GlyAsnGlyGlyHisGlyAlaThrAsnThrAlaAlaThrAlaThr 592

Qy    850 --------------------------------GGTCAGGGCGGTAATGCTGGCACGGGT 876
                                          |||   |||||||||   |||||||||
Db    593 GlyGlyAlaGlyGlyAlaGlyGlyIleLeuPheGlyThrGlyGlyAsnGlyGlyThrGly 612

Qy    877 CTTGACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTAC 936
                  |||   :::   |||      |||:::                |||
Db    613 ------GlyIleAlaThrGlyAlaGlyGlyIleGlyGlyAlaGlyGlyAlaGlyGlyVal 630

Qy    937 CAGCAGTTAGGTAACGCCGTGGGTACCGGT-----------------ATCGGTATGAAA 978
                  |||     :::   |||||||||                    ||||||:::
Db    631 SerLeuLeuIleGlySerGlyGlyThrGlyGlyAsnGlyGlyAsnSerIleGlyVal--- 649

Qy    979 GCGGGCATTCAGGCGCTGAATGATATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTC 1038
            |||||||||                      |||
Db    650 AlaGlyIleGlyGlyAlaGlyGlyArgGly---------------------------- 659

Qy   1039 GTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCT 1098
                     ||||||   :::
Db    660 ---------GlyAspAlaGlyLeu----------------------------------- 664

Qy   1099 GAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGAC 1158
               :::||||||                  |||   |||
Db    665 ---LeuPheGly-----------AlaAlaGlyThrGlyGly------------------ 673

Qy   1159 AAATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATG 1218

Db    673 ----------------------------------------------------------- 673

Qy   1219 GAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGGCCCATGGCGGGTGATACCGGCAAC 1278
                     :::   |||   |||      :::          :::   |||      ||||||
Db    674 ---------HisGlyAlaAlaGlyGlyValProAlaGlyValGlyGlyAlaGlyGlyAsn 690

Qy   1279 GGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCGCTGGGTATTGATGCCATGATGGCC 1338
            |||   |||   |||   |||      |||   :::   |||   :::|||
Db    691 GlyGlyLeuPheAlaAsnGlyGlyAlaGlyGlyAlaGlyGlyPheAsnAla---AlaGly 709

Qy   1339 GGTGATGCCATTAACAATATGGCACTTGGCAAGCTGGGC 1377
            |||:::      |||            |||      |||
Db    710 GlyAsnGlyGlyAsnGlyGlyLeuPheGlyThrGlyGly 722

RESULT 9
YZ08_MYCTU
ID   YZ08_MYCTU     STANDARD;      PRT;  1901 AA.
AC   O53553;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv3508 precursor.
GN   RV3508 OR MTV023.15.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
```

```
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL022022; CAA17745.1; -.
DR   TubercuList; Rv3508; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1901       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                                RV3508.
SQ   SEQUENCE   1901 AA;  147627 MW;  C7B1923D5D0146CD CRC64;

Alignment Scores:
Pred. No.:              5.18e-08      Length:         1901
Score:                  240.00        Matches:        131
Percent Similarity:     34.97%        Conservative:   26
Best Local Similarity:  29.18%        Mismatches:     180
Query Match:            9.48%         Indels:         112
DB:                     1             Gaps:           19

SEQ12-SEQ4 (1-1390) x YZ08_MYCTU (1-1901)

Qy     199 GGAGCGTCAACGATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTG 255
Db    1133 GlyAlaSerThrThrSerIleAsnAlaAsnGlyGlyAlaGlyGlyAsnGlyGlyThrGly 1152

Qy     256 GGTACCAGTCGCCAGAATGCTGGGTTGGTGGCAATTCTGCACTGGGGCTGGGC------ 309
Db    1153 Gly-----------LysGlyGlyAlaGlyGlyAlaGlyThrLeuGlyValGlyGlySer 1168

Qy     309 ------------------------------------------------------------ 309

Db    1169 GlyGlyThrGlyGlyAspGlyGlyAspAlaGlySerGlyGlyGlyGlyGlyPheGlyGly 1188

Qy     310 ------------------------------GGCGGTAATCAAAATGATACCGTCAATCAG 339
Db    1189 AlaAlaGlyLysAlaGlyGlyGlyGlyAsnGlyGlyValGlyGlyAspGlyGlyGluGly 1208

Qy     340 CTGGCTGGCTTACTCACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTG 399
Db    1209 AlaSerGlyLeuGlyLeuGlyLeuSerGlyPheAspGlyGlyGlnGlyGlyGlnGlyGly 1228

Qy     400 ATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTG--- 456
Db    1229 AlaGlyGlySerAlaGlyAlaGlyGlyIleAsnGlyAlaGlyGlyAlaGlyGlyThrGly 1248

Qy     457 GGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC-----------------GGT 498
Db    1249 GlyAlaGlyGlyAspGlyAlaProAlaThrLeuIleGlyGlyProAspGlyGlyAspGly 1268

Qy     499 TCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCCCCG 558
Db    1269 GlyGlnGlyGlyIleGlyGlyAspGlyGlyAsn--AlaGlyPheGlyAlaGlyValProG 1288

Qy     559 CT-----------GGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCC 606
Db    1288 lyAspGlyGlyAspGlyGlyAsnAlaGlyPheGlyAlaGlyValPro-GlyAspGlyGly 1307

Qy     607 ACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATG 666
Db    1308 IleGlyGlyThrGlyGlyAlaGlyGlyAlaGly-------------------------- 1318

Qy     667 TTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAA---------GATGGCACCCAG 717
Db    1319 -----------------GlyAlaGlyAlaAspGlyAspProSerIleAspGlyGlyGln 1332

Qy     718 GGCAGTTCCTCTGGG---GGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
Db    1333 GlyGlyAlaGlyGlyHisGlyGlyGlnGlyGlyLysGlyGlyLeuAsnSer------Thr 1350

Qy     775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAAT---GGTCTGAGCCAGCTCCTTGGC 831
Db    1351 GlyLeuAlaSerAlaAlaSerGlyAspGlyGlyAsnGlyGlyAlaGlyGlyAlaGlyGly 1370

Qy     832 AACGGGGGACTGGGAGGTGGT---CAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCG 888
Db    1371 AsnGlyGlyAspGlyAspGlyPheIleGlyGlySerGlyGlyThrGlyGlyThrGlyGly 1390

Qy     889 TCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGT 948
Db    1391 AspAlaGlyValGlyGlyLeuAlaAsnThrGlyGlyThrAla--------------Gly 1405

Qy     949 AACGCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGT 1008
Db    1406 AsnAlaGlyIleGlyGlyAlaGlyGlyArgGlyGlyAspGlyGlyAlaGlyAspSerGly 1425

Qy    1009 ACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAG 1068
Db    1426 --AlaLeuSerGln----------------------------------AspGlyAsnG 1433

Qy    1069 GAAATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAA 1128
Db    1433 lyPheAlaGlyGlyGlnGly-----------------GlyGlnGlyGlyValGlyGlyA 1447

Qy    1129 GGCCCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCA 1188
Db    1447 snAlaGlyAlaGlyGlyIleAsnGlyAlaGlyGlyThrGlyGlyThrGlyGlyAlaGlyG 1467

Qy    1189 GATGACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATC 1248
Db    1467 lyAspGlyGlnAsnGlyThrThr------GlyValAlaSerGluGlyGlyAlaGlyGlyG 1485

Qy    1249 AAAAGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGT 1308
Db    1485 ln---------Gly-GlyAspGlyGlyGlnGlyGlyIle---------GlyGlyAlaGly 1498

Qy    1309 TCTTCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGC 1368
Db    1499 GlyAsnAlaGlyPheGlyAlaGlyValProGlyAspGly-----------GlyIleGly 1514

Qy    1369 AAGCTGGGCGCGGCT 1383
Db    1515 GlyThrGlyGlyAla 1519

RESULT 10
YI40_MYCTU
ID   YI40_MYCTU     STANDARD;      PRT;   515 AA.
AC   Q50594;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv1840c.
GN   RV1840C OR MT1888 OR MTCY1A11.04 OR MTCY359.33.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
```

```
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z83859; CAB06114.1; -.
DR   EMBL; AE007047; AAK46159.1; -.
DR   TIGR; MT1888; -.
DR   TubercuList; Rv1840c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      1     21       POTENTIAL.
FT   TRANSMEM    165    185       POTENTIAL.
FT   TRANSMEM    199    219       POTENTIAL.
SQ   SEQUENCE   515 AA;  43916 MW;  2EDC8B6C6D28B7F3 CRC64;

Alignment Scores:
Pred. No.:              6.26e-08      Length:         515
Score:                  238.50        Matches:        109
Percent Similarity:     33.08%        Conservative:   23
Best Local Similarity:  27.32%        Mismatches:     128
Query Match:            9.42%         Indels:         139
DB:                     1             Gaps:           15

SEQ12-SEQ4 (1-1390) x YI40_MYCTU (1-515)

Qy     226 GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGT 285
           |||||||||||||||:::  ||||||:::|||              |||   |||:::|||
Db     148 GlyGlyAlaGlyGlyAspAlaGlyLeuIleGlyAsnGly---GlyAsnGlyGlyIleGly 166

Qy     286 GGCAATTCTGCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCT 345
           |||      |||  ||||||  ||||||                               |||
Db     167 GlyProGlyAlaThrGlyLeuAlaGlyGly---------------------------Ala 177

Qy     346 GGCTTACTCACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGC 405
           |||   :::   |||::::::                  ||||||   |||    :::|||
Db     178 GlyGlyValGlyGlyLeuLeuPheGlyAspGlyGlyAsnGlyGlyAlaGlyGlyLeuGly 197

Qy     406 GGTGGCTTAGGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGA 465
           |||      |||      |||   |||::::||||||   ||||||            |||
Db     198 ThrGlyProValGlyAlaThrGly---GlyIleGlyGlyProGlyGlyAlaAlaValGly 216

Qy     466 CTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGC 525
           |||                              |||   |||         |||     |||
Db     217 LeuPheGlyHisGlyGlyAlaGlyGlyAlaGlyGlyLeuGlyLysAlaGlyPheAlaGly 236

Qy     526 GGCAACAATACCACTTCAACAACAAATTCCCCGCTG------GACCAGGCGCTGGGTATT 579
           |||         |||        |||          |||        :::          |||
Db     237 GlyAlaGlyGlyThrGlyGlyThrGlyGlyLeuLeuTyrGlyAsnGlyGlyAsnGlyGly 256

Qy     580 AACTCAACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGC 639
           |||      |||         |||     :::|||   |||:::
Db     257 AsnValProSerGlyAlaAlaAspGlyGlyAlaGlyGlyAspAla--------------- 271

Qy     640 GACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGAT 699
                                                         |||   |||:::
Db     272 ---------------------------------------------ArgLeuIleGlyAsn 276

Qy     700 GGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAG 759
           |||   ||||||                    |||   |||     ||||||
Db     277 GlyGlyAspGly------------GlySerValGlyAlaAlaProThrGly--------- 289

Qy     760 AACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGC 819
                                                    :::|||||||||
Db     290 -----------------------------------------IleGlyAsnGly------ 293

Qy     820 CAGCTCCTTGGCAACGGGGGACTGGGAGGT---------GGTCAGGGCGGTAATGCTGGC 870
                    ||||||||||||   ||||||          |||   ||||||:::   |||
Db     294 ---------GlyAsnGlyGlyAsnGlyGlyTrpLeuTyrGlyAspGlyGlySerGlyGly 310

Qy     871 ACGGGTCTTGACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTG 930
           :::   |||   |||   |||   ||||||   |||
Db     311 SerThrLeuGlnGlyPheSerAspGlyGlyThrGly------------------------ 322

Qy     931 GACTACCAGCAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAG 990
                          |||||||||         |||   |||            |||
Db     323 ---------------GlyAsnAlaGlyMetPheGlyAspGlyGlyAsnGlyGlyPheSer 337

Qy     991 GCGCTGAATGATATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGC 1050
              :::      |||         :::                   |||   |||
Db     338 PhePheAspGlyAsnGlyGlyAspGlyGlyThrGlyGlyThrLeuIleGlyAsnGlyGly 357

Qy    1051 GATCGGGCGATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGC 1110
           |||
Db     358 Asp--------------------------------------------------------- 358

Qy    1111 AAGCCGCAGTACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCA 1170
                                   |||      |||:::||||||
Db     359 --------------------GlyGlyAsnSerValGlnThrAsp--------------- 366

Qy    1171 AAAGCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAAC 1230

Db     366 ------------------------------------------------------------ 366

Qy    1231 AAAGCCAAGGGCATGATCAAAAGGCCCATGGCGGGTGATACCGGCAAC------------ 1278
                    |||   :::   |||         ||||||   ||||||
Db     367 ---------GlyPheLeu---ArgGlyHisGlyGlyAspGlyGlyAsnAlaValGlyLeu 382

Qy    1279 ---GGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCGCTGGGTATTGATGCCATGATG 1335
              ||||||   |||   |||||||||   :::      |||:::   |||
Db     383 IleGlyAsnGlyGlyAlaGlyGlyAlaGlySerAlaGlyThrGlyValPheAlaProGly 402

Qy    1336 GCCGGTGATGCCATTAACAAT-----------ATGGCACTTGGCAAGCTGGGCGCG 1380
              |||         |||              :::   :::|||      ||||||
Db     403 GlyGlySerGlyGlyAsnGlyGlyAsnGlyAlaLeuLeuValGlyAsnGlyGlyAla 421

RESULT 11
EXLP_TOBAC
ID   EXLP_TOBAC     STANDARD;      PRT;   426 AA.
AC   Q03211;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Pistil-specific extensin-like protein precursor (PELP).
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=cv. Petit Havana; TISSUE=Pistil;
RX   MEDLINE=93005740; PubMed=1392607;
RA   Goldman M.H., Pezzotti M., Seurinck J., Mariani C.;
RT   "Developmental expression of tobacco pistil-specific genes encoding
RT   novel extensin-like proteins.";
RL   Plant Cell 4:1041-1051(1992).
CC   -!- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER
CC       PISTIL DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING
CC       FLOWER DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE
CC       AFTER POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER
CC       POLLINATION.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z14019; CAA78397.1; -.
DR   PIR; JQ1696; JQ1696.
DR   InterPro; IPR000419; Pollen_Ole_e_I.
DR   Pfam; PF01190; Pollen_Ole_e_I; 1.
KW   Structural protein; Signal; Repeat; Glycoprotein.
FT   SIGNAL        1     23
FT   CHAIN        24    426       PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN.
FT   DOMAIN       69    182       4 X 5 AA REPEATS OF S-P(4).
FT   REPEAT       69     73       1.
FT   REPEAT       76     80       2.
FT   REPEAT       83     87       3.
FT   REPEAT      178    182       4.
FT   CARBOHYD    310    310       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   426 AA;  44278 MW;  51A495CC94017812 CRC64;

Alignment Scores:
Pred. No.:              8.95e-08      Length:          426
Score:                  236.00        Matches:         88
Percent Similarity:     35.86%        Conservative:    21
Best Local Similarity:  28.95%        Mismatches:      85
Query Match:            9.10%         Indels:          110
DB:                     1             Gaps:            17

SEQ12-SEQ4 (1-1390) x EXLP_TOBAC (1-426)

Qy   1025 GAACTGTGCCTGTGCGTACCGATATCATTCAGCGCCTGAATGCCCGCTTTCATACCGATA 966
            ||||||   |||    :::|||:::    |||            ||||||    |||||||:::
Db     29 GluLeuTrpLeuGluLeuProLeuProPheAspTrp---ProProAlaGluIleProLeu 47

Qy    965 CCG---GTACCCACGGCGTTACCTAACTGCTGGTAGTCCACCGGCCCGCTCAGGTTTTGC 909
            |||      :::|||:::                                ||||||      |||
Db     48 ProAspIleProSerProPhe-----------------AspGlyPro---ThrPheVal 60

Qy    908 AGCCCTTTGCCGCCCAGCGACGAACCGTCAAGACCCGTGCCAGCA---TTACCGCCCTGA 852
                |||    |||||||||          ||||||    |||    |||:::        ||||||
Db     61 LeuPro---ProProSerProLeuProSerProProProProSerProSerProProPro 79

Qy    851 CCACCTCCCAGTCCCCCGTTGCCAAGGAGC------------------------------ 822
            |||    ||||||||||||     |||
Db     80 ProSerProSerProProProProSerThrIleProLeuIleProProPheThrGlyGly 99

Qy    821 TGGCTCAGACCATTACCC------ATCAGGCCCGACAGCGCATCAGTGACTCCTTTTTTA 768
            :::|||    |||||||||            ||||||    |||     :::    |||     :::
Db    100 PheLeuProProLeuProGlySerLysLeuProAspPheAlaGlyLeuLeuProLeuIle 119

Qy    767 TAGGCGTTCTGCTCG---CCTTCGGTCGGCTGCTTGCCCCCAGAGGAACTGCCCTGGGTG 711
                                |||    :::|||        ||||||          |||
Db    120 ProAsnLeuProAspValProProIleGlyGlyGlyProProValAsnGlnPro---Lys 138

Qy    710 CCATCTTGCCCATCACCAAACAGGCTTTGCATTATCTCGCTGAACATCTTCAGCAGCTGC 651
            ||||||    |||||||||
Db    139 ProSerSerProSerPro------------------------------------------ 144

Qy    650 TGCATCGGGTCGCTGGAGTCTGAGGTGGAATCTGTGCCGGAGGTGGAATCGTCGTTTTGG 591

Db    144 ------------------------------------------------------------ 144

Qy    590 GACGTTGAGTTAATACCCAGCGCCTGGTCCAGCGGGGAATTTGTTGTTGAAGTGGTATTG 531
                                                                  :::|||
Db    145 ------------------------------------------------LeuValLys 147

Qy    530 TTGCCGCCTTTCGAGCCCAGCGTGTTCAGCGAACCGCCTAACATATCGTTCAGCGCGTTC 471
                ||||||      ||||||          |||
Db    148 ProProProProProProSerProCysLysProSerProPro------------------ 161

Qy    470 GACAGTCCTTCGCCCAGGCACCTGAG---CCACCCAAGCCATTACCTAAGCCACCGCCT 414
            |||    :::    :::||||||:::    ||||||          |||    |||||||||
Db    162 AspGlnSerAlaLysGlnProProGlnProProProAlaLysGlnProSerProProPro 181

Qy    413 AAGCCACCGCCCATCAGCCCACCACCGCCCATCATGCTCATCATCATCATCATGCCGGTG 354
               |||||||||:::       |||    |||                      |||
Db    182 ProProProProValLysAlaProSerPro-------------------SerProAla 194

Qy    353 AGTAAGCCAGCCAGCTGATTGACGGTATCATTTTGATTACCGCCGCCCAGCCCCAGTGCA 294
               :::|||                                   |||||||||    |||
Db    195 LysGlnPro-----------------------------ProProProProProVal 204

Qy    293 GAATTGCCA---CCCAACCCAGCATTCTGG--------------------------CGA 264
            :::    |||    |||:::||||||
Db    205 LysAlaProSerProSerProAlaThrGlnProProThrLysGlnProProProProPro 224

Qy    263 CTGGTACCCAGCAACCCGTTATTTCCGCCCGCACCGCCGATAGAAATTTGCATCGTTGAC 204
                       :::||||||    ||||||    ||||||:::              |||
Db    225 ArgAlaLysLysSerProLeuLeuProProProProProValAlaTyrProProValMet 244

Qy    203 GCTCCCAGCCCA 192
               |||||||||
Db    245 ThrProSerPro 248

RESULT 12
WA22_MYCTU
ID   WA22_MYCTU    STANDARD;     PRT;   914 AA.
AC   O06794;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   WAG22 antigen precursor.
GN   WAG22 OR RV1759C OR MT1807 OR MTCY28.25C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
```

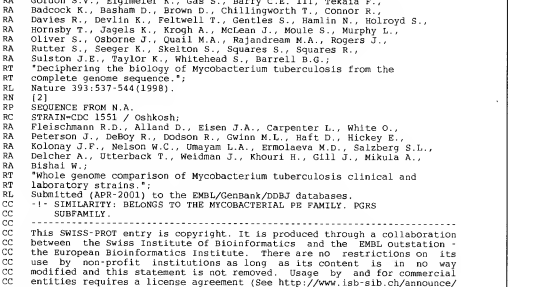
```
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT IN POSITION 85.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z95890; CAB09322.1; -.
DR   EMBL; AE007040; -; NOT_ANNOTATED_CDS.
DR   TIGR; MT1807; -.
DR   TubercuList; Rv1759c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
KW   Antigen; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    914       WAG22 ANTIGEN.
SQ   SEQUENCE   914 AA;  74354 MW;  F6953C3DBE8E6AC8 CRC64;

Alignment Scores:
Pred. No.:              9.78e-08      Length:         914
Score:                  235.50        Matches:        120
Percent Similarity:     33.91%        Conservative:   38
Best Local Similarity:  25.75%        Mismatches:     173
Query Match:            9.30%         Indels:         135
DB:                     1             Gaps:           20

SEQ12-SEQ4 (1-1390) x WA22_MYCTU (1-914)

Qy      193 GGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAATAACGGG--- 249
                |||   |||   |||   |||              ||||||||||||||      |||
Db      186 GlyAlaGlyGlySerAlaMetLeuPheGlyAlaGlyGlyAlaGlyGlyAlaGlyGlyAla 205

Qy      250 ---------TTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTGCACTG 300
                      |||:::|||                |||   ||||||||:::    :::
Db      206 AlaThrSerLeuValGlyGlyIleGlyGlyThrGlyGlyThrGlyGlyAsnAlaGlyMet 225

Qy      301 --------------------GGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAG 339
                                 |||      ||||||         |||
Db      226 LeuAlaGlyAlaAlaGlyAlaGlyGlyAlaGlyGlyPheSerPheSerThrAlaGlyGly 245

Qy      340 CTGGCTGGCTTA---CTCACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGG 396
                  |||          |||    :::                :::||||||   |||
Db      246 AlaGlyGlyAlaGlyGlyAlaGlyGlyLeuPheThrThrGlyGlyValGlyGlyAlaGly 265

Qy      397 CTGATGGGC-----------------------GGTGGCTTAGGCGGTGGCTTAGGT--- 429
                  |||                         ||||||  ||||||      |||
Db      266 GlyGlnGlyHisThrGlyGlyAlaGlyGlyAlaGlyGlyAlaGlyGlyLeuPheGlyAla 285

Qy      430 AATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAA------------GGACTGTCGAACGCG 477
                |||:::|||||||:::||||||   |||:::               |||       |||
Db      286 GlyGlyMetGlyGlyAlaGlyGlyPheGlyAspHisGlyThrLeuGlyThrGlyGlyAla 305

Qy      478 CTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACC 537
                  |||      ||||||   |||            |||      ||||||
Db      306 GlyGlyAspGlyGlyGlyGlyGlyLeuPheGlyAlaGlyGlyAspGlyGlyAlaGlyGly 325

Qy      538 ACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAAC 597
                :::      |||                      |||    |||:::   :::   :::
Db      326 SerGlyLeuThrThrGlyGlyAla---AlaGlyAsnGlyGlyAsnAlaGlyThrLeuSer 344

Qy      598 GACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTG 657
                   :::      |||          |||         :::
Db      345 LeuGlyAlaAlaGlyGlyAlaGlyGlyThrGlyGlyAlaGlyGly--------------- 359

Qy      658 CTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACCCAG 717
                                           ::::::||||||   |||:::      |||
Db      360 --------------------------ThrValPheGlyGlyGlyLysGlyGlyAlaGly 370

Qy      718 GGCAGT-----------------------TCCTCTGGGGGCAAGCAGCCGACCGAAGGC 753
                |||:::                        |||||||||            |||   |||
Db      371 GlyAlaGlyGlyAsnAlaGlyMetLeuPheGlySerGlyGlyGlyGlyGlyThrGlyGly 390

Qy      754 GAGCAGAACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGT 813
                                     |||                  |||    :::|||
Db      391 -----------------PheGlyPheAlaAlaGlyGlyGlnGlyGlyValGlySer 404

Qy      814 CTGAGCCAGCTCCTTGGCAACGGGGACTGGGAGGTGGTCAGGGC--------------- 858
                         |||   |||:::||||||   ||||||      |||
Db      405 AlaGlyMetLeuSerGlySerGlyGlySerGlyGlyAlaGlyGlySerGlyGlyProAla 424

Qy      859 -----------------------------------------------------GGTAATGCT 867
                                                                  ||||||
Db      425 GlyThrAlaAlaGlyGlyAlaGlyGlyAlaGlyGlyAlaProGlyLeuIleGlyAsnGly 444

Qy      868 GGCACGGGTCTTGACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCG 927
                |||   |||   :::|||         ||||||   |||          :::|||
Db      445 GlyAsnGlyGlyAsnGlyGlyGluSerGlyGlyThrGlyGlyValGlyGlyAlaGly--- 463

Qy      928 GTGGACTACCAGCAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATT 987
                              ||||||||||||      |||   |||    :::   ||||||
Db      464 -----------------GlyAsnAlaValLeuIleGlyAsnGlyGlyGluGlyGlyIle 477

Qy      988 CAGGCG-----------------------------------------CTGAATGATATC 1005
                  |||                                          |||
Db      478 GlyAlaLeuAlaGlyLysSerGlyPheGlyGlyPheGlyGlyLeuLeuLeuGlyAlaAsp 497

Qy     1006 GGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCG 1065
                |||   :::         ||||||      |||                  :::
Db      498 GlyTyrAsnAlaProGluSerThrSerProTrpHisAsn--------------LeuGln 512

Qy     1066 AAGGAAATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTT---GGCAAGCCGCAGTAC 1122
                ::::::|||         |||:::::::::        |||         |||:::|||
Db      513 GlnAspIleLeuSerPheIleAsnGluProThrGluAlaLeuThrGlyArgProLeuIle 532

Qy     1123 CAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGC 1182
                   |||
Db      533 GlyAsnGly--------------------------------------------------- 535

Qy     1183 AAGCCAGATGACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGC 1242
                            |||      ||||||   :::   :::               |||   |||
Db      536 -----------AspSerGlyThrProGlyThrGlyAspAsp---GlyGlyAlaGlyGly 550

Qy     1243 ATGATCAAAAGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCC 1302
                                 :::   |||:::   |||||||||         |||   |||
Db      551 Trp-----------LeuPheGlyAsnGlyGlyAsnGlyGlyAlaGlyAlaAlaGlyThr 566

Qy     1303 GGTGGTTCTTCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCA 1362
                ||||||:::   |||                  ||||||   |||          :::
Db      567 AsnGlySerAlaGlyGly--------------AlaGlyGlyAla---GlyGlyIleLeu 580

Qy     1363 CTTGGCAAGCTGGGCGCG 1380
                |||      ||||||
Db      581 PheGlyThrGlyGlyAla 586

RESULT 13
YP91_MYCTU
ID   YP91_MYCTU     STANDARD;      PRT;   543 AA.
AC   Q50630;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv2591.
GN   RV2591 OR MT2668.1 OR MTCY227.10C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licence@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z77724; CAB01283.1; -.
DR   EMBL; AE007100; AAK46982.1; ALT_INIT.
DR   TIGR; MT2668.1; -.
DR   TubercuList; Rv2591; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   543 AA;  46287 MW;  59730339E5D2DF59 CRC64;
```

Alignment Scores:

Pred. No.:	1.04e-07	Length:	543
Score:	235.00	Matches:	113
Percent Similarity:	36.80%	Conservative:	32
Best Local Similarity:	28.68%	Mismatches:	122
Query Match:	9.28%	Indels:	127
DB:	1	Gaps:	19

SEQ12-SEQ4 (1-1390) x YP91_MYCTU (1-543)




```
RESULT 14
GRP_ARATH
ID   GRP_ARATH       STANDARD;       PRT;   338 AA.
AC   P27483;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   Glycine-rich cell wall structural protein precursor.
OS   Arabidopsis thaliana (Mouse-ear cress).
```

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=92003708; PubMed=1912511;
RA   Quigley F., Villiot M.L., Mache R.;
RT   "Nucleotide sequence and expression of a novel glycine-rich protein
RT   gene from Arabidopsis thaliana.";
RL   Plant Mol. Biol. 17:949-952(1991).
CC   -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X58338; CAA41249.1; -.
DR   PIR; S17732; KNMU.
KW   Cell wall; Structural protein; Repeat; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    338       GLYCINE-RICH CELL WALL STRUCTURAL
FT                                PROTEIN.
FT   DOMAIN       21    338       GLY-RICH.
SQ   SEQUENCE   338 AA;  23891 MW;  046A6E8C1A4E89EB CRC64;

Alignment Scores:
Pred. No.:              1.28e-07      Length:         338
Score:                  233.50        Matches:        80
Percent Similarity:     35.07%        Conservative:   14
Best Local Similarity:  29.85%        Mismatches:     77
Query Match:            9.22%         Indels:         97
DB:                     1             Gaps:           10

SEQ12-SEQ4 (1-1390) x GRP_ARATH (1-338)

Qy    193 GGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAATAACGGGTTG 252
            |||||||||                       ||||||||||||||       |||:::
Db     59 GlyLeuGly-----------------------GlyGlyAlaGlyGlyGlyGlyGlyIle 70

Qy    253 CTGGGTACCAGTCGCCAGAATGCT------GGGTTGGGTGGCAATTCTGCACTGGGGCTG 306
              |||   :::        |||      |||||||||||    :::      ||||||
Db     71 GlyGlyGlyAlaGlyGlyGlyAlaGlyGlyGlyLeuGlyGlyGlyAlaGlyGlyGlyLeu 90

Qy    307 GGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATG 366
            ||||||||:::                        |||    :::    |||
Db     91 GlyGlyGlyHis-----------------------GlyGlyGlyIleGlyGly------ 100

Qy    367 ATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTA 426
                            |||   ||||||      ||||||||||||||||||||
Db    101 -----------------GlyAlaGlyGlyGlyAlaGlyGlyGlyLeuGlyGlyGlyHis 114

Qy    427 GGTAATGGCTTGGGTGGC------TCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTG 480
            |||   |||:::||||||            ||||||   |||   ||||||         :::
Db    115 GlyGlyGlyIleGlyGlyGlyAlaGlyGlyGlySerGlyGlyGlyLeuGlyGlyGlyIle 134

Qy    481 AACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACT 540
                        ||||||                 |||      ||||||    :::
Db    135 GlyGlyGlyAlaGlyGlyGlyAlaGlyGlyGlyGlyGlyLeuGlyGlyGlyHis------ 152

Qy    541 TCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGAC 600

Db    152 ------------------------------------------------------------ 152

Qy    601 GATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTG 660

Db    152 ------------------------------------------------------------ 152

Qy    661 AAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACCCAGGGC 720
                                           |||   |||      |||      |||
Db    153 -------------------------------GlyGlyGlyIleGlyGlyGlyAlaGly 161

Qy    721 AGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAGGAGTC 780
                :::||||||               |||            :::      |||:::
Db    162 GlyGlyAlaGlyGly---------------GlyLeuGlyGlyGlyHisGlyGlyGlyIle 176

Qy    781 ACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAACGGGGA 840
                         |||      |||    ||||||                 ||||||
Db    177 GlyGlyGlyAlaGlyGlyGlySerGlyGlyGlyLeu---------------GlyGlyGly 191

Qy    841 CTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTGGGCGGC 900
            :::|||||||||   ||||||    ||||||    |||      |||   :::   ||||||
Db    192 IleGlyGlyGlyAlaGlyGlyGlyAlaGlyGlyGly------GlyGlyAlaGlyGlyGly 209

Qy    901 AAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCCGTGGGT 960
               ||||||          |||                         |||        |||
Db    210 GlyGlyLeuGlyGlyGlyHisGly-----------GlyGlyPheGlyGlyGlyAlaGly 225

Qy    961 ACCGGTATCGGTATGAAAGCGGGC 984
               |||:::|||      ||||||
Db    226 GlyGlyLeuGlyGlyGlyAlaGly 233

RESULT 15
YD25_MYCTU
ID   YD25_MYCTU     STANDARD;      PRT;   603 AA.
AC   Q10637;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv1325c precursor.
GN   RV1325C OR MT1367 OR MTCY130.10C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z73902; CAA98089.1; -.
DR   EMBL; AE007010; AAK45630.1; ALT_INIT.
DR   TIGR; MT1367; -.
DR   TubercuList; Rv1325c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Repeat; Signal; Complete proteome.
FT   SIGNAL         1     30       POTENTIAL.
FT   CHAIN         31    603       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                                 RV1325C.
FT   DOMAIN       114    603       GLY-RICH.
FT   CONFLICT     132    132       G -> V (IN REF. 2).
FT   CONFLICT     135    135       G -> D (IN REF. 2).
FT   CONFLICT     337    337       G -> D (IN REF. 2).
FT   CONFLICT     508    508       L -> F (IN REF. 2).
SQ   SEQUENCE   603 AA;  49575 MW;  4F9BCB82B07AE964 CRC64;

Alignment Scores:
Pred. No.:                1.39e-07      Length:         603
Score:                    233.00        Matches:        120
Percent Similarity:       33.77%        Conservative:   33
Best Local Similarity:    26.49%        Mismatches:     139
Query Match:              9.20%         Indels:         161
DB:                       1             Gaps:           25

SEQ12-SEQ4 (1-1390) x YD25_MYCTU (1-603)

Qy     132 GGTCGGCAGGGTACGTTTGAATTATTCATAAGAGGAATACGTT---------------AT 176
           ||||||   |||   ||||||:::               |||                 ::
Db     243 GlyArgSerGlyLeuPheGluValLeuAlaAlaGly-GlyAlaGlyGlyThrGlyGlyLe 262

Qy     177 GAGTCTGAAT---------------ACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTC 221
           :|||:::|||                ||| ||| |||                  ||
Db     262 uSerValAsnGlyGlyThrGlyGlyThrGlyGlyThrGlyGlyGlyGlyGlyLeuPheSe 282

Qy     222 TATCGGCGGTGCGGGCGGAAATAAC------GGTTGCTGGGTACCAGTCGCCAGAATGC 275
           |  ||||||||||||||              |||:::  ||||:::::        |||
Db     282 rAsnGlyGlyAlaGlyGlyAlaGlyGlyPheGlyValSerGlySerAlaGlyGlyAsnGl 302

Qy     276 TGGGTTGGGTGGCAATTCTGCACTG---------------GGGCTGGGCGGCGGTAATCA 320
            |||  ||||||:::      :::                |||   ||||||
Db     302 yGlyThrGlyGlyAspGlyGlyIlePheThrGlyAsnGlyGlyThrGlyGlyThrGlyGl 322

Qy     321 AAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATGAGCAT 380
                |||   |||||||||  |||               |||          .
Db     322 yThrGlyThrGlyAsnGlnLeuValGlyGlyGluGlyGly------------------- 335

Qy     381 GATGGGCGGTGGTGGG---------------CTGATGGGC---GGTGGCTTAGGCGGT-- 420
                ||||||  |||                |||  |||   |||||||:::||||||
Db     336 -AlaGlyGlyAlaGlyGlyAsnAlaGlyIleLeuPheGlyAlaGlyGlyIleGlyGlyTh 355

Qy     421 ----------GGCTTAGGT-----------AATGGCTTGGGTGGCTCAGGTGGCCTGGG 458
                     |||||||||              |||  ||||||  |||||||:::||
Db     355 rGlyGlyThrGlyLeuGlyAlaProAspProGlyGlyThrGlyGlyLysGlyGlyValGl 375

Qy     459 CGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTC 518
           |   |||:::       |||                 |||:::|||          |||
Db     375 y---GlyIleGlyGlyAla-----------------GlyAlaLeuPheGlyProGlyGl 388

Qy     519 GAAAGGCGGCAACAATACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTAT 578
                ||||||                                            |||
Db     388 yAlaGlyGlyThrGly---------------------------------GlyPheGlyAl 397

Qy     579 TAACTCAACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAG 638
             :::|||      |||                    |||:::      :::      :::::
Db     397 aSerSerAlaAspGlnMetAlaGlyGlyIleGlyGlySerGlyGlySerGlyGlyAlaAl 417

Qy     639 CGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGA 698
            :                                                  |||   |||||
Db     417 a--------------------------------------------LysLeuIleGlyAs 422

Qy     699 TGGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCA 758
           ||||     |||            :::||||||
Db     422 pGlyGlyAlaGly-----------GlyThrGlyGly---------------------- 430

Qy     759 GAACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAG 818
                                   |||:::::: |||     |||:::|||
Db     431 ------------------------AspSerValArgGlyAlaAlaGlySerGly----- 440

Qy     819 CCAGCTCCTTGGCAACGGGGGACTGGGA--------------GGTGGTCAGGGCGGTAA 863
                     |||   ||||||  |||                ||||||   ||||||
Db     441 ---------GlyThrGlyGlyThrGlyGlyLeuIleGlyAspGlyGlyAlaGlyGlyAl 457

Qy     864 TGCTGGCACGGGTCTTGACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGG 923
               |||||||||::::::      |||:::||||||  |||
Db     457 aGlyGlyThrGlyIleGluPheGlySerValGlyGlyAlaGlyGlyAlaGly-------- 474

Qy     924 GCCGGTGGACTACCAGCAGTTAGGTAACGCCGTGGGT---ACCGGTATCGGTATGAAAGC 980
                                 |||||||||  |||   :::|||   |||
Db     475 ---------------------GlyAsnAlaAlaGlyLeuSerGlyAlaGlyGlyAlaGl 487

Qy     981 GGGCATTCAGGCGCTGAATGATATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTCGT 1040
             |||           :::
Db     487 yGlyAlaGlyGlyPheGlyGlu------------------------------------- 494

Qy    1041 CAATAAAGGCGATCGGGCGATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCTGA 1100
                   ||||||                 |||
Db     495 -ThrAlaGlyAspGlyGlyAlaGlyGlyAsnAlaGlyLeuLeu---------------- 508

Qy    1101 GGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGACAA 1160
                                        |||   |||
Db     509 ------------------------AsnGlyAspGly-----------------GlyAl 514

Qy    1161 ATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATGGA 1220
                 |||      |||                |||   |||
Db     514 aGlyGlyAlaGlyGlyLeuGlyIleAlaGlyAspGlyGly------------------- 527

Qy    1221 GCAGTTCAACAAAGCCAAGGGCATGATCAAAAGGCCCATGGCGGGTGATACCGGCAACGG 1280
                 |||     ||||||     |||       |||   |||:::    |||:::||
Db     528 -------AsnGlyGlyLysGlyGly---LysAlaGlyMetValGlyAsnGlyGlyAspGl 544

Qy    1281 CAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCGCTGGGTATTGATGCCATGATGGCCGG 1340
           |              ||||||||||||:::|||:::         :::      :::  ||
Db     544 y--------------GlyAlaGlyGlyAlaSerValValAlaAsnGlyGlyValGlyGl 559

Qy    1341 TGATGCCATTAACAATATGGCACTTGGCAAGCTGGGC 1377
           |         |||         :::|||    |||
Db     559 ySerGlyGlyAsnAlaThrLeuIleGlyAsnGlyGly 571

Search completed: January 14, 2003, 17:09:14
Job time : 27 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run On: January 14, 2003, 17:06:02 ; Search time 27 Seconds
(without alignments)
9898.285 Million cell updates/sec

Title: SEQ12-SEQ4
Perfect score: 2532
Sequence: 1 atgggatttttctcttttc.....gctgggcgcggcttaagctt 1390

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Pgapop 6.0 , Pgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/KUBELIK693/runat_14012003_091819_7089/app_query.fasta_1.1543
-DB=PIR_73 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cd1 -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=KUBELIK693_&CGN_1_1_37_0runat_14012003_091819_7089 -NCPU=6 -ICPU=3
-NO_XLFXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -PGAPOP=6 -PGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2050	81.0	403	2	T08471	harpin - Erwinia a
2	264.5	10.4	1489	2	D70807	hypothetical glyci
3	259	10.2	1660	2	A70869	hypothetical glyci
4	257.5	10.2	2174	2	E95965	hypothetical glyci
5	255.5	10.1	1381	2	E70806	hypothetical glyci
6	255	10.1	591	2	B70523	hypothetical glyci
7	255	10.1	1538	2	H70846	hypothetical glyci
c 8	254	9.8	620	2	S06733	hydroxyproline-ric
9	254	10.0	1079	2	B70807	hypothetical glyci
c 10	253.5	9.8	464	2	S22697	extensin - Volvox
11	253.5	10.0	588	2	F70971	hypothetical glyci
c 12	253.5	9.8	744	2	B86255	hypothetical prote
c 13	251	9.7	368	2	C29356	hydroxyproline-ric
14	249	9.8	731	2	C70974	hypothetical glyci
15	248	9.8	463	2	B70893	hypothetical glyci
16	248	9.8	714	2	A70807	hypothetical glyci
17	248	9.8	778	2	F70963	hypothetical glyci
c 18	245	9.5	489	2	T11622	extensin class 1 p
19	245	9.7	923	2	E70820	hypothetical glyci
20	241.5	9.5	957	2	D70835	hypothetical glyci
c 21	241	9.3	1188	2	S49915	extensin-like prot
22	240.5	9.5	615	2	H70589	hypothetical glyci
23	240	9.5	1901	2	F70806	hypothetical glyci
c 24	239	9.2	760	2	T06291	extensin homolog T
25	239	9.4	882	2	B70812	hypothetical glyci
26	238.5	9.4	515	2	H70663	hypothetical glyci
c 27	237.5	9.2	280	2	T11671	extensin-like prot
c 28	237.5	9.2	1611	2	T38236	hypothetical prote
29	236.5	9.3	853	2	A70896	hypothetical glyci
c 30	236	9.1	426	2	JQ1696	pistil extensin-li
c 31	236	9.1	5262	2	T03454	ALR protein - huma
32	235.5	9.3	914	2	H70987	hypothetical glyci
33	235	9.3	439	2	D70954	hypothetical glyci
34	235	9.3	543	2	F70726	hypothetical glyci
35	234.5	9.3	1329	2	E70917	hypothetical glyci
c 36	234	9.0	599	2	T10798	pherophorin-S - Vo
37	233.5	9.2	338	1	KNMU	glycine-rich cell
38	233.5	9.2	434	2	E70768	hypothetical glyci
39	233	9.2	603	2	A70770	hypothetical glyci
40	232.5	9.2	1306	2	A70934	hypothetical glyci
41	232	9.2	801	2	F70824	hypothetical glyci
42	232	9.2	837	2	E70835	hypothetical glyci
c 43	230.5	8.9	839	2	T04859	extensin homolog F
44	230	9.1	667	2	A70893	hypothetical glyci
45	229	9.0	491	2	D70916	hypothetical glyci

ALIGNMENTS

RESULT 1
T08471
harpin - Erwinia amylovora
C;Species: Erwinia amylovora
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08471
R;Laby, R.J.; Kim, J.F.; Beer, S.V.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z16433
A;Accession: T08471
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-403 <LAB>
A;Cross-references: EMBL:M92994; NID:g4309677; PID:g4309678
A;Experimental source: strain 321
C;Genetics:
A;Gene: hrpN

Alignment Scores:
Pred. No.: 2.79e-147 Length: 403
Score: 2050.00 Matches: 399
Percent Similarity: 99.01% Conservative: 0
Best Local Similarity: 99.01% Mismatches: 4
Query Match: 80.96% Indels: 0
DB: 2 Gaps: 0

SEQ12-SEQ4 (1-1390) x T08471 (1-403)

```
Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy      355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy      415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy      475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy      535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy      595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy      655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy      715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy      775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy      835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy      895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy      955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy     1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      281 SerAspSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy     1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy     1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy     1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysSer 360

Qy     1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 AlaMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy     1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy     1375 GGCGCGGCT 1383
            |||||||||
Db      401 GlyAlaAla 403
```

RESULT 2
D70807
hypothetical glycine-rich protein Rv3514 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: D70807
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: D70807
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1489 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17751.1; PID:g292
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3514
C;Superfamily: collagen alpha 1(IV) chain

```
Alignment Scores:
Pred. No.:               4.76e-12     Length:         1489
Score:                   264.50       Matches:        116
Percent Similarity:      32.13%       Conservative:   27
Best Local Similarity:   26.07%       Mismatches:     149
Query Match:             10.45%       Indels:         153
DB:                      2            Gaps:           16

SEQ12-SEQ4 (1-1390) x D70807 (1-1489)

Qy      181 CTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATT-------------------- 219
              ||||||||  ||||||    ::::::|||
Db      955 LeuAsnThrAspGlyLeuSerSerAlaThrSerGlyThrGlyGlyThrGlyGlyThrGly 974

Qy      220 ---------------------------------------------------TCTATCGGC 228
                                                                   |||
Db      975 GlyLysGlyGlyThrGlyGlyAlaGlyAspAspSerAlaGlyGlyThrGlyGlyThrGly 994

Qy      229 GGTGCGGGCGGAAATAACGGGTTG--------------CTGGGTACCAGTCGCCAGAAT 273
            ||||||||||||||    |||                 |||    :::        |||
Db      995 GlyAlaGlyGlyAsnAlaGlyAlaGlyGlyLeuAlaAsnThrGlyGlyThrAlaGlyAsn 1014

Qy      274 GCTGGGTTGGGTGGCAATTCTGCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTC 333
            |||||:::||||||:::        |||    |||    |||   |||   |||:::
Db     1015 AlaGlyIleGlyGlyAspGlyGlyGlnGly---GlyAsnGlyGlyGlnGlyAspSerGly 1033

Qy      334 AATCAGCTG-----------------GCTGGCTTACTCACCGGCATGATGATGATGATG 375
            :::   |||                 ||||||        |||
Db     1034 SerGlyLeuGlyGlyGlnProGlyPheAlaGlyGlyAlaGlyGlyLysGlyGlyAlaGly 1053

Qy      376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                     |||    ||||||      |||    |||    ||||||    |||    |||
Db     1054 GlySerSerGlyAlaGlyGlyThrAsnGlySerGlyGlyAlaGlyGlyAlaGly---Gly 1072

Qy      436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
            ||||||:::||||||    |||         ||||||         |||          |||
Db     1073 GlnGlyGlyAlaGlyGlyAlaGlyIleSerPheSerAsnGlySerAsnGlyGlyThrGly 1092

Qy      496 GGTTCG------CTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACA 549
            |||:::      :::          |||      ||||||||
Db     1093 GlyThrGlyGlyValGlyGlyThrGlyGlyAspGlyGlyAsn----------------- 1106

Qy      550 AATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACC 609
                                  |||       :::                       |||
Db     1107 --------------------AlaGlyThrGlyAlaGlyAspProGlyLysGlyGlyThr 1119

Qy      610 TCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTC 669
              ||||||      |||      |||
Db     1120 GlyGlyThrGlyGlyThrGlyGlySerGly----------------------------- 1129
```

A;Gene: Rv2490c
C;Superfamily: collagen alpha 1(IV) chain

Alignment Scores:
Pred. No.: 1.26e-11 Length: 1660
Score: 259.00 Matches: 119
Percent Similarity: 33.41% Conservative: 22
Best Local Similarity: 28.20% Mismatches: 144
Query Match: 10.23% Indels: 137
DB: 2 Gaps: 19

SEQ12-SEQ4 (1-1390) x A70869 (1-1660)

```
Qy   1057 GCGAT-------------------------------------GGCGAAGGAAATCGGT 1077
          |||||                                     |||:::|||
Db    380 lyAspGlyGlyAlaGlyGlyGlyGlyGlyLeuGlyGlyAlaProGlyAspGlyGlyAlaG 400

Qy   1078 CAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGT 1137
                :::|||          |||   |||   ||||||              |||
Db    400 lyGlyAsnGly-----------GlySerTrpLeuAlaAla--------------GlyA 411

Qy   1138 CAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGAC 1197
             ||||||
Db    411 spGlyGlyAlaGly-------------------------------------------- 415

Qy   1198 GGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGGCCC 1257
                                                                     |
Db    416 ---------------------------------------------------------GlyH 417

Qy   1258 ATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCGCTG 1317
           ||||| ||||||    |||    |||           |||   |||||||||||:::|||
Db    417 isGly-GlyAspProGlyLeuGlyGly------AlaGlyGlyAlaGlyGlyAlaSerGly 434

Qy   1318 GGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTGGGC 1377
           |||     |||      ||||||         |||   :::|||  |||      |||
Db    435 GlyAlaGlyAla---ArgAlaGly------AlaAsnGlyLeuAlaAlaGlyAsnAspGly 451
```

RESULT 4
E95965
hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) maga
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: E95965
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: E95965
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2174 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:g15140875; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMb21548
A;Genome: plasmid

Alignment Scores:

Pred. No.:	1.7e-11	Length:	2174
Score:	257.50	Matches:	125
Percent Similarity:	38.58%	Conservative:	49
Best Local Similarity:	27.72%	Mismatches:	163
Query Match:	10.17%	Indels:	114
DB:	2	Gaps:	25

SEQ12-SEQ4 (1-1390) x E95965 (1-2174)

```
Qy    133 GTCGGCAGGGTACGTTTGAATTATTCATAAGAGGAATACGTTATGAGTCTGAATACAAGT 192
          ||||||   :::::::::|||                  |||    |||:::    |||
Db    343 ValGlyAlaIleGlnIleAsnVal------------TyrGlyGlySerIleGluThrAsp 358

Qy    193 GGGCTGGGAGCG---TCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAATAACGGG 249
          |||:::   |||          :::     ||||||||||||   |||||||||   |||
Db    359 GlyIleGlnAlaTyrGlyLeuLeuGlyGlySerIleGlyGlyGlnGlyGlyAsnGlyGly 378

Qy    250 ------------TTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTGCA 297
                      ::::::|||||||||              |||   ||||||||||:::
Db    379 AspAspThrAlaValValGlyThrSer---GlyGlyGlyGlyPheGlyGlyAsnAlaGly 397

Qy    298 CTGGGGCTGGGCGGC------------GGTAATCAAAATGATACCGTCAATCAGCTGGCT 345
             |||:::|||             |||::::::        |||              :::
Db    398 ---GlyValGlyAlaTyrValThrThrGlySerLysIleThrThrThrGlyAspPheSer 416

Qy    346 GGCTTACTCACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGC 405
             :::|||             :::      :::||||||||||||      |||
Db    417 AlaAlaValThr--------------LeuHisSerIleGlyGlyGlyGlyGlyThrGly 431

Qy    406 GGTGGC------------TTAGGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGC 453
           |||               ||||||||||||   |||||||||      |||      |||
Db    432 GlyAspPheThrAspValLeuGlyGlyGlyAlaGlyAsnGlyGlyAsnGlyGlyAsnGly 451

Qy    454 CTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTG 513
                      ::::::||||||                  ||||||:::::::::|||
Db    452 AsnThrAlaThrIleAsnAsnAla---------------GlyGlyThrIleSerThrSer 466

Qy    514 GGC--------------------------TCGAAAGGCGGCAACAATACCACTTCAACA 546
           |||                          |||   ||||||:::        |||
Db    467 GlyGlyHisSerTyrGlyIleLeuValGlnSerIleGlyGlySerGlyGlyThrGlyGly 486

Qy    547 ACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCC 606
                      |||         ||||||   :::                :::
Db    487 IleAlaGluGlyLeuThrLeuGluLeuGlyGlyAspGly---------GlyGluGlyGly 503

Qy    607 ACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATG 666
           ::::::|||      |||      ::::::
Db    504 SerAlaGlyAlaAlaSerValGlnAsnThrGlyAlaIleThrThrAspGlyTyrSerAla 523

Qy    667 TTCAGCGAGATAATGCAAAGCCTGTTT---GGTGATGGGCAAGATGGCACCCAGGGC--- 720
                   :::    ||||||:::       |||   |||         ||||||   |||
Db    524 HisGlyIleLeuAlaGlnSerIleSerGlyGlyGlyGlyAlaAlaGlyThrAlaGlyGly 543

Qy    721 ------------AGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTAT 768
                            :::||||||                  |||
Db    544 IleLeuSerIleGlyGlyThrGlyGlyAsnAlaAsnTyrAlaGlyAsnAlaGluValGln 563

Qy    769 AAAAAAGGAGTCACT------------GATGCGCTGTCGGGCCTGATGGGTAATGGTCTG 816
            |||                       ||||||   |||::::::               :::
Db    564 SerSerGlyAlaIleSerThrGlyGlyAspAlaAlaIleGlyIleLeuAlaGlnSerIle 583

Qy    817 AGCCAGCTCCTTGGCAACGGGGGACTGGGAGGTGGTCAG-----------------GGC 858
                      |||    ||||||   |||||||||                     |||
Db    584 Gly---------GlyGlyGlyGlySerGlyGlyGlyAlaLysGlyIleAlaThrValGly 600

Qy    859 GGTAATGCTGGCACGGGT-------------------------CTTGACGGTTCGTCG 891
          |||:::   |||   |||                          |||:::|||   |||
Db    601 GlySerGlyGlyAlaGlyGlyTyrGlyThrTyrAlaSerValTrpLeuAsnGlyGlySer 620

Qy    892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
          :::      :::|||          ::::::      :::      |||   :::|||
Db    621 IleAlaThrArgGly------GluMetAlaHisGlyIleMetAlaGlnSerIleGlyGly 638

Qy    952 GCCGTGGGTACCGGT---------ATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGAT 1002
                |||   |||          |||   |||         ||||||   ||||||
Db    639 GlyGlyGlyAsnGlyGlyAsnValIleAspMetSerValGlyIleProAlaLeuGly--- 657

Qy   1003 ATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATG 1062
          :::|||                                           |||   |||     |
Db    658 ValGly----------------------------------------Gly--SerAlaThrG 664

Qy   1063 GCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTAC 1122
          ||    |||:::                                 |||   |||
Db    664 lyGlyGlySer---------------------------GlyGlyTrpValCysValThrA 675

Qy   1123 CAGAAAGGCCCGGG---------TCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAA 1173
            |||   ||||||          |||   |||          |||   :::        :
Db    675 snGluAsnProGlyGlyAspCysSerThrGlySerGlyAlaProGlnAlaValAlaIleA 695

Qy   1174 GCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAA 1233
          ::|||      |||                               |||   :::    ||||
Db    695 laThrAlaGlyAla--------------------GlyAlaValGlyIleLeuAlaGlnS 708
```

```
Qy   1234 GC-----------CAAGGGCATGATCAAAAGGCCCATGGCGGGTGATACCGGCAACGGC 1281
            ||            |||:::          |||   |||  ||||||   |||   |||
Db    708 erIleGlyGlyGlyGlyGlyAsnGlyGlyAsnAlaThrGly-GlyAspAlaGlyPheGly 727

Qy   1282 AACCTGCAGGCACGCGGTGCCGGTGGT 1308
          :::   |||      |||   ||||||
Db    728 SerPheGlnIleGlyGlyGlyGlyGly 736
```

RESULT 5
E70806
hypothetical glycine-rich protein Rv3507 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: E70806
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: E70806
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1381 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17744.1; PID:g292444
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3507
C;Superfamily: collagen alpha 1(IV) chain

Alignment Scores:

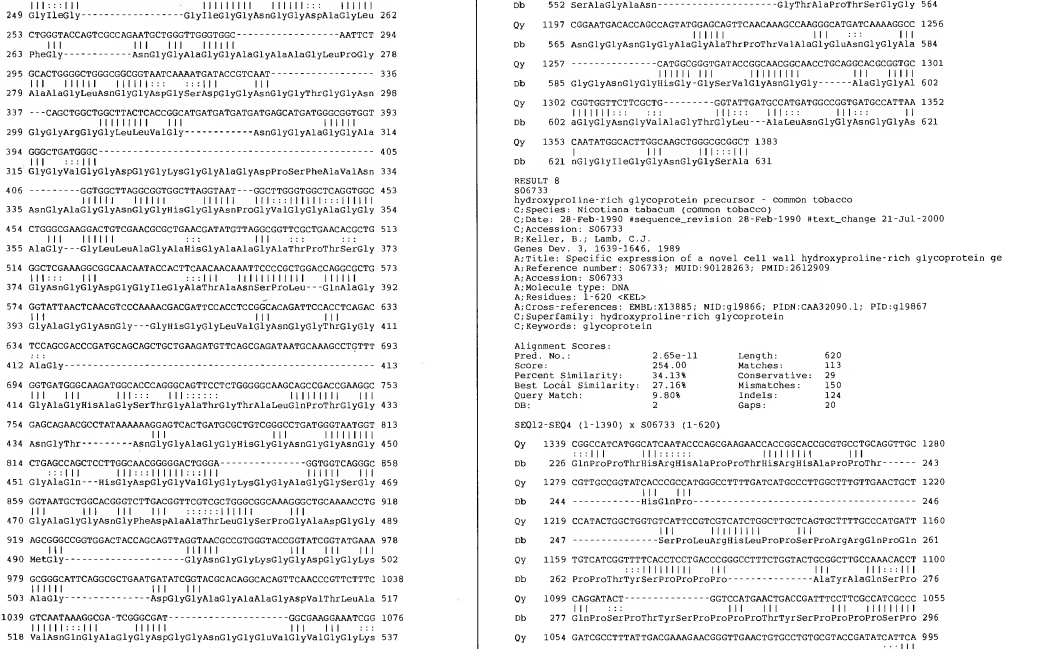
Pred. No.:	2.26e-11	Length:	1381
Score:	255.50	Matches:	128
Percent Similarity:	35.04%	Conservative:	29
Best Local Similarity:	28.57%	Mismatches:	187
Query Match:	10.09%	Indels:	105
DB:	2	Gaps:	21

SEQ12-SEQ4 (1-1390) x E70806 (1-1381)

```
Qy    184 AATACAAGTGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAAT 243
          |||      |||   |||                  |||||||||||||||||
Db    479 AsnProGlyGlyLysGlyGlyAspGly---------GlyIleGlyGlyAlaGlyGlyAla 495

Qy    244 AACGGGTTGCTGGGTACCAGTCGC-----------CAGAATGCTGGGTTGGGTGGCAAT 291
             |||      ||||||                   |||    |||    ||||||:::
Db    496 GlyGlyAlaAlaGlyThrGlyAsnGlyGlyHisAlaGlyAsnThrGlyAspGlyGlyAsp 515

Qy    292 TCTGCACTGGGGCTGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTA 351
                |||      |||||||||              ||||||
Db    516 GlyGlyThrGlyGlyAsnGlyGlyAsnGlyThrGlyGlyValAsnGlyAlaAspAsnThr 535

Qy    352 CTCACCGGCATGATGATGATGATGAGCATGATGGGCGGT-----------GGTGGCTG 399
           |||                           ||||||          ||||||
Db    536 LeuAsn--------------ProAspThrProGlyGlyAlaGlyGluProGlyGlyAla 550

Qy    400 ATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGC 459
             |||   |||      ||||||        |||   ||||||:::||||||   |||
Db    551 GlyGlyAlaGlyGlyAlaGlyGlyAlaAlaGlyGlyProGlyGlyThrGlyGlyThrGly 570

Qy    460 GAAGGACTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCG 519
                |||      |||          ||||||:::              |||
Db    571 GlyAsnGlyGlyAsnGlyGlyAsnGlyGlyAsnGlyGlyAsnGlyGlyAsn---GlyGly 589

Qy    520 AAAGGCGGCAACAATACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATT 579
             ||||||||       ::::::||||||:::|||                :::|||
Db    590 AsnGlyGlyAsnAlaGlyAsnAsnSerThrAsnAlaPro------------ValGlyGly 605

Qy    580 AACTCAACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGC 639
                :::      |||          |||          ::::::
Db    606 GluGlyGlyAlaGlyGlyAspGlyGlyAlaGlyGlyAlaGlyGlyAlaAlaAsnGlyGly 625

Qy    640 GACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGAT 699
                                          :::      |||      |||
Db    626 -----------------------------ThrAlaGlySerGlnGlyThrGlyGlyVal 635

Qy    700 GGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAG 759
           |||   ||||||      |||   :::   ||||||   :::         |||
Db    636 GlyGlyAspGlyGlyAlaGlyGlyAsnGlyGlyGlyGlyLys------AlaGlyThrGly 653

Qy    760 AACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGC 819
           |||:::         ||||||   |||      :::|||      |||
Db    654 AsnSerGlyAsnPheGlyVal---AspGlyGluAlaGlyPheSerGlyGlyAla------ 670

Qy    820 CAGCTCCTTGGCAACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTT 879
                   ||||||||||||:::||||||      |||   |||   |||||||||
Db    671 ------GlyGlyAsnGlyGlyValGlyGlyAlaAlaGlyAlaAsnGlyGlyThrGlyGly 688

Qy    880 GACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAG 939
             |||   :::   |||      |||         :::   |||
Db    689 SerGlyGlyAsnGlyGlyAspGlyGlyAlaGlyGlyIleGlyGlyAlaGly--------- 705

Qy    940 CAGTTAGGTAACGCCGTG-------------------------GGTACCGGTATC--- 969
                ||||||   :::                          |||||||||
Db    706 ------GlyAsnGlyIleProGlyThrGlyThrGluProAlaGlyGlyThrGlyAlaLys 723

Qy    970 GGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCACAGGCACAGTTCAACC 1029
           |||         |||         |||   :::               :::
Db    724 GlyGlyAspGlyGlyAspGlyGlyAlaGlyGlyAlaGlyGlyAsnAlaGlyGlyAlaGly 743

Qy   1030 CGTTCTTTCGTCAATAAAGGCGA-------TCGGGCGATGGCGAAGGAAATCGGTCAGTT 1082
                   |||   |||         :::|||         ||||||      :::
Db    744 GlyGlnGlyGlyAsnAlaGlyGlnGlyGlyAlaGlyGlyAlaGlyGlyAsnAlaValIle 763

Qy   1083 CATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGTCAGGA 1142
           |||                ||||||   :::|||         :::      |||||||||
Db    764 ProGlyAsp---------GlyValGlyLysAlaProHisGlyAspAlaGlyGlySerGly 780

Qy   1143 GGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGACGGAAT 1202
           |||:::                ||||||      :::
Db    781 GlyAspGly--------------GlyLysGlyGlyGlnGlyGly---------SerGly 792

Qy   1203 GACACCAGCCAGTATGGAGCA----------------------GTTCAACAAAGCCAA 1238
             |||         ||||||                              |||
Db    793 GlyThrGlyGlySerGlyAlaProIleGlyGlyGlyAlaGlyGlyThrGlyGlySerGly 812

Qy   1239 GGGCATGATCAAAAG----------------------------------------- 1253
           ||||||      |||
Db    813 GlyHisAlaGlyLysGlyGlyAlaGlyGlyIleGlyAlaGlnGlyThrThrIleThrVal 832

Qy   1254 -----------GCCCATGGCGGGTGATACCGGCAAC---GGCAACCTGCAGGCACGCGG 1298
                       ::: |||||||||   ||||||   ||||||      |||   ||
Db    833 ProGlyAsnGlyGlyAsn-AlaGlyAspGlyGlyAsnGlyGlyAsnAlaGlyAlaGlyGl 852

Qy   1299 TGCCGGTGGTTCTTCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATAT 1358
           |   |||||||||                          ||||||      |||
Db    852 yAsnGlyGlySer--------------------------GlyAspPheGlyGlyAsnTh 863

Qy   1359 GGCACTTGGCAAGCTGGGCGCG 1380
                |||      |||:::
Db    863 rThrSerGlyAlaSerGlySer 870
```

RESULT 6
B70523
hypothetical glycine-rich protein Rv0297 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jan-2000
C;Accession: B70523
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

```
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: B70523
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-591 <COL>
A;Cross-references: GB:Z96800; GB:AL123456; NID:g3261800; PIDN:CAB09596.1; PID:e321655;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0297
C;Superfamily: unassigned collagens

Alignment Scores:
Pred. No.:              2.21e-11      Length:         591
Score:                  255.00        Matches:        123
Percent Similarity:     32.03%        Conservative:   24
Best Local Similarity:  26.80%        Mismatches:     168
Query Match:            10.07%        Indels:         144
DB:                     2             Gaps:           20

SEQ12-SEQ4 (1-1390) x B70523 (1-591)

Qy    193 GGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAATAACGGGTTG 252
          |||   |||   |||                         ||||||||||||||      |||
Db    141 GlyAsnGlyGlySerGlyAlaProGlyGlnAlaGlyGlyAlaGlyGlyAlaAlaGlyPhe 160

Qy    253 CTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTGCACTGGGGCTGGGCGGC 312
             |||                ||| |||  ||||||:::          |||     |||
Db    161 PheGly-----------AsnGlyGlyAsnGlyGlyAspGlyGlyAlaGlyAlaAsnGly 176

Qy    313 GGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATG 372
          |||                          ||||||       |||
Db    177 Gly-------------------------AlaGlyGlyThrAlaGlyTrpPhePheGly 187

Qy    373 ATGAGCATGATGGGCGGTGGTGGGCTGATGGGC--------------GGTGGCTTAGGC 417
                      ||||||  |||   :::|||               |||||||||||
Db    188 PheGlyGlyAsnGlyGlyAlaGlyGlyIleGlyValAlaGlyIleAsnGlyGlyLeuGly 207

Qy    418 GGTGGCTTAGGTAAT------------------------------GGCTTGGGTGGCTCA 447
          |||      |||:::                             |||   ||||||
Db    208 GlyAlaGlyGlyAspGlyGlyAsnAlaGlyPhePheGlyAsnGlyGlyAsnGlyGlyMet 227

Qy    448 GGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTA-------------- 492
          ||||||  |||        |||||:::|||      |||
Db    228 GlyGlyAlaGlyAlaAlaGlyValAsnAlaValAsnProGlyLeuAlaThrProValThr 247

Qy    493 --------------GGCGGTTCGCTGAACACGCTGGGCTCGAAAGGC------GGCAAC 531
                        |||    ||||||   :::|||        |||       |||
Db    248 ProAlaAlaAsnGlyGlyAsnGlyLeuAsnLeuValGlyValProGlyThrAlaGlyGly 267

Qy    532 AATACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCC 591
                                |||   :::   ||||||   |||
Db    268 GlyAlaAspGlyAlaAsnGlySerAlaIleGlyGlnAlaGlyGlyAlaGlyGlyAspGly 287

Qy    592 CAAAACGACGATTCCACCTCCGGC--------------ACAGATTCCACCTCAGACTCC 636
            |||     ::::::   |||                    |||            :::
Db    288 GlyAsnAlaSerThrSerGlyGlyIleGlyIleAlaGlnThrGlyGlyAlaGlyGlyAla 307

Qy    637 AGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGT 696
                                                                  |||
Db    308 Gly------------------------------------------GlyAlaGlyGly 312

Qy    697 GATGGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAG 756
          ||||||      |||        |||    :::    |||          :::   |||
Db    313 AspGlyAlaProGlyGlyAsnGlyGlyAsnGlyGlySerValGluHisThr--------- 329

Qy    757 CAGAACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTG 816
                                |||   |||   :::   |||       |||||||||
Db    330 -----------------GlyAlaThrGlySer---SerAlaSerGlyGlyAsnGlyAla 342

Qy    817 AGCCAGCTCCTTGGCAACGGGGGACTGGGA----------------------------- 846
          :::             ||||||||||||:::|||
Db    343 ThrGly------GlyAsnGlyGlyValGlyAlaProGlyGlyAlaGlyGlyAsnGlyGly 360

Qy    847 ----------------------------------GGTGGTCAGGGCGGTAATGCTGGC 870
                                            ||||||:::|||||||||    |||
Db    361 HisValSerGlyGlySerValAsnThrAlaGlyAlaGlyGlyLysGlyGlyAsnGlyGly 380

Qy    871 ACGGGTCTTGACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTG 930
          ||||||       |||         ||||||    |||         |||||||||||||||
Db    381 ThrGlyGlyAlaGlyGlyPro---GlyGlyHisGlyGlySerValLeuSerGlyProVal 399

Qy    931 GACTACCAGCAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAG 990
                         |||   |||   |||   |||             ||||||:::
Db    400 GlyAspSerGlyAsnGlyGlyAlaGlyGlyAspGly---------GlyAlaGlyValSer 416

Qy    991 GCGCTGAATGATATCGGTACGCACAGGCACAGTTCAACCCGTTCT-----------TTC 1038
          |||    :::      ||||||                                   :::
Db    417 AlaThrAspIleAlaGlyThrGlyGlyArgGlyGlyAsnGlyGlyHisGlyGlyLeuTrp 436

Qy   1039 GTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCT 1098
          :::       |||               |||       :::|||
Db    437 IleGlyAsnGlyGlyAspGlyGlyAlaGlyGlyValGly-------------------- 449

Qy   1099 GAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGAC 1158
                                        |||   |||
Db    450 ----------------------------GlyValGly-----------------Gly 453

Qy   1159 AAATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATG 1218
                   |||   |||:::               ||||||
Db    454 AlaGlyAlaAlaGlyAlaIleGlyGlyHisGlyGlyAspGly----------------- 467

Qy   1219 GAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGGCCCATG-----------GCGGGT 1266
                                  |||    :::    |||:::          ||||||
Db    468 -------------------GlySerValAsnThrProIleGlyGlySerGluAlaGly 480

Qy   1267 GATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCGCTGGGTATTGAT 1326
          |||   |||   |||   |||         |||   |||              |||
Db    481 AspGlyGlyLysGlyGlyLeuGlyGlyAspGlyGlyGlyArgGlyIlePheGly------ 498

Qy   1327 GCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTGGGCGCGGCT 1383
                   ||||||   |||                :::|||      |||   |||
Db    499 GlnPheGlyAlaGlyGlyAlaGlyGlyAlaGlyGlyValGlyGlyAlaGlyGlyAla 517

RESULT 7
H70846
hypothetical glycine-rich protein Rv3345c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: H70846
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70846
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1538 <COL>
A;Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAA17117.1; PID:g326
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3345c
C;Superfamily: collagen alpha 1(IV) chain

Alignment Scores:
Pred. No.:              2.51e-11      Length:         1538
Score:                  255.00        Matches:        133
Percent Similarity:     36.14%        Conservative:   30
Best Local Similarity:  29.49%        Mismatches:     167
Query Match:            10.07%        Indels:         122
DB:                     2             Gaps:           23

SEQ12-SEQ4 (1-1390) x H70846 (1-1538)
```



RESULT 8
S06733
hydroxyproline-rich glycoprotein precursor - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000
C;Accession: S06733
R;Keller, B.; Lamb, C.J.
Genes Dev. 3, 1639-1646, 1989
A;Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein ge
A;Reference number: S06733; MUID:90128263; PMID:2612909
A;Accession: S06733
A;Molecule type: DNA
A;Residues: 1-620 <KEL>
A;Cross-references: EMBL:X13885; NID:g19866; PIDN:CAA32090.1; PID:g19867
C;Superfamily: hydroxyproline-rich glycoprotein
C;Keywords: glycoprotein

Alignment Scores:
Pred. No.: 2.65e-11 Length: 620
Score: 254.00 Matches: 113
Percent Similarity: 34.13% Conservative: 29
Best Local Similarity: 27.16% Mismatches: 150
Query Match: 9.80% Indels: 124
DB: 2 Gaps: 20

SEQ12-SEQ4 (1-1390) x S06733 (1-620)

```
Qy     934 AGTCCACCGGCCCGCTCAGGTTTTGC--------------AGCCCTTTGCCGCCCAGCG 890
              |||      |||       :::                 ||||||  |||||||:::
Db     318 --SerProProProProAlaTyrSerProProProThrTyrSerProProProProThrT 337

Qy     889 ACGAACCG--------------------TCAAGACCCGTGCCAGCATTACCGCCCTGAC 851
               |||                       |||   |||   |||         |||   |
Db     337 yrLeuProLeuProSerSerProIleTyrSerProProProValTyrSerProProP 357

Qy     850 CACCTCCC------AGTCCCCCGTTGCCAAGGAGCTGGCTCAGACCATTACCCATCAGGC 797
           ||||||||      |||||||||   |||                |||   |||      |
Db     357 roProProSerTyrSerProProProProThrTyrLeuProProProProProSerSerP 377

Qy     796 CCGACAGCGCATCAGTGACTCCTTTTTTATAGGCGTTCTGCTCGCCTTCGGTCGGCTGCT 737
            ||        |||    :::|||                   |||:::
Db     377 roProProProSerPheSerPro--------------ProProProThrTyrGluGlnS 392

Qy     736 TGCCCCCAGAGGAACTGCCCTGGGTGCCATCTTGCCCATCACCAAACAGGCTTTGCATTA 677
              ||||||          :::   |||     |||:::|||
Db     392 erProProProProProAlaTyrSerProProLeuProAlaProProThrTyr------- 409

Qy     676 TCTCGCTGAACATCTTCAGCAGCTGCTGCATCGGGTCGCTGGAGTCTGAGGTGGAATCTG 617
                                              |||                  |||
Db     410 ---------------------------------SerProProProProThrTyrSerP 418

Qy     616 TGCCGGAGGTGGAATCGTCGTTTTGGGACGTTGAGTTAATACCCAGCGCCTGGTCCAGCG 557
              |||              ::::::               :::|||      :::|||
Db     418 roPro------ProProThrTyrAlaGlnProProProLeuProProThrTyrSer---- 434

Qy     556 GGGAATTTGTTGTTGAAGTGGTATTGTTGCCGCCTTTCGAGCCCAGCGTGTTCAGCGAAC 497
                                          ||||||     |||      :::|||    |
Db     435 ----------------------------ProPro------ProProAlaTyrSerProP 443

Qy     496 CGCCTAACATATCGTTCAGC--------------------------------GCGTTCG 470
            |||||      :::::|||                                 |||:::
Db     443 roProProProThrTyrSerProProProProThrTyrSerProProProProAlaTyrA 463

Qy     469 ACAGTCCTTCGCCCAGGCCACCT------GAGCCACCCAAGCCATTACCTAAGCCACCGC 416
               |||   |||   ||||||           ||||||    |||         |||||||
Db     463 laGlnProProProProProProThrTyrSerProProProProAlaTyrSerProProP 483

Qy     415 CTAAGCCACCGCCCATCAGCCCACCACCGCCCATCATGCTCATCATCATCATCATGCCGG 356
            ||       |||    ||||||||||||||          :::       :::   |||
Db     483 roProSerProIleTyrSerProProProPro------GlnValGlnProLeuProProT 501

Qy     355 TGAGTAAGCCAGCCAGCTGATTGACGGTATCATTTTGATTACCGCCGCCCAGCCCCAGTG 296
                  |||               :::          |||||||||||
Db     501 hrPheSerProProProProArgArgIleHis------LeuProProProProHisArgG 519

Qy     295 CAGAATTGCCACCCAACCCAGCATTCTGGCGACTGGTACCCAGCAACCCGTTATTTCCGC 236
                  ||||||   |||    :::    :::          ||||||    |||    |||    |
Db     519 lnProArgProProThrProThrTyrGlyGln---ProProSerProProThrPheSerP 538

Qy     235 CCGCACCGCCGATAGAAATTTGCATCGTTGACGCTCCCAGCCCA 192
            ||    ||||||    :::|||              :::|||    |||
Db     538 roProProProArgGlnIle---------HisSerProProPro 549
```

RESULT 9
B70807
hypothetical glycine-rich protein Rv3512 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: B70807
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: B70807
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1079 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17749.1; PID:g292
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3512
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

```
Alignment Scores:
Pred. No.:                 2.85e-11        Length:          1079
Score:                     254.00          Matches:         123
Percent Similarity:        36.16%          Conservative:    35
Best Local Similarity:     28.15%          Mismatches:      166
Query Match:               10.03%          Indels:          113
DB:                        2               Gaps:            18
```

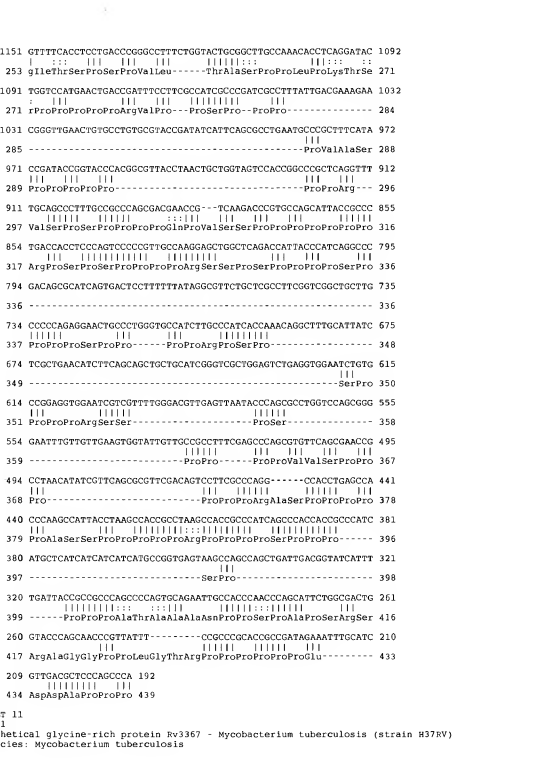
SEQ12-SEQ4 (1-1390) x B70807 (1-1079)

```
Qy     187 ACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAATAAC 246
           |||   |||   |||   :::                :::   |||||||||||||||
Db     636 ThrGlyGlyAlaGlyGlyAsnGlyGlyGlyGlyAlaAsnGlyGlyAlaGlyGlyAlaGly 655

Qy     247 GGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTGCACTGGGGCTG 306
           |||      |||   :::      |||   |||   ||||||::::::
Db     656 GlySerGlyGlyGlyThrGlyGlyAsnGlyGlyAlaGlyGlyAspAlaGlyAspAlaGly 675

Qy     307 GGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATG 366
              |||||||||   |||   |||   |||           |||
Db     676 AsnGlyGlyAsnGlyAsnGlyThrGlyAsnGlyGlyAsnGly----------------- 689

Qy     367 ATGATGATGAGCATGATGGGCGGTGGTGGGCTG------ATGGGCGGTGGCTTAGGCGGT 420
                                         |||   ||||||:::      |||||||||      |||
Db     690 -----------------GlyAsnGlyGlyIleAlaGlyMetGlyGlyAsnGlyGlyAla 703

Qy     421 GGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTG 480
           |||   |||:::|||   ||||||:::||||||   |||               |||
Db     704 GlyThrGlySerGlyAsnGlyGlyAsnGlyGlySerGly---------GlyAsnGlyGly 720

Qy     481 AACGATATGTTAGGCGGTTCGCTGAACACG---------CTGGGCTCGAAAGGCGGCAAC 531
           |||      :::||||||:::         |||              |||      |||||||||
Db     721 AsnAlaGlyMetGlyGlyAsnSerGlyThrGlySerGlyAspGlyGlyAlaGlyGlyAsn 740

Qy     532 AATACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACG--- 588
                 :::   |||                              |||:::         |||
Db     741 GlyGlyAlaAlaGlyThrGlyGlyThrGlyGlyAspGly---GlyLeuThrGlyThrGly 759

Qy     589 --------------------------TCCCAAAACGACGATTCCACCTCCGGCACAGAT 621
                                           :::         :::|||      |||
Db     760 GlyThrGlyGlySerGlyGlyThrGlyGlyAspGlyGlyAsnGlyGlyAsnGlyAlaAsp 779

Qy     622 TCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATG 681
           :::|||::::::                                          :::
Db     780 AsnThrAlaAsn------------------------------------------MetThr 785

Qy     682 CAAAGCCTGTTTGGTGATGGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAG 741
                          |||||||||   :::|||         |||         ||||||
Db     786 AlaGlnAlaGlyGlyAspGlyGlyAsnGlyGlyAspGlyGlyPheGlyGlyGlyAlaGly 805

Qy     742 CCGACCGAAGGCGAGCAGAACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTG 801
                   |||         |||         |||   |||                |||
Db     806 AlaGlyGlyGlyGlyLeuThrAlaGlyAlaAsnGly---ThrGlyGlyGlnGlyGlyAla 824

Qy     802 ATGGGTAATGGTCTGAGCCAGCTCCTTGGCAACGGGGGACTGGGAGGT----------- 849
              |||:::|||                  |||||||||      :::||||||
Db     825 GlyGlyAspGly--------------GlyAsnGlyAlaIleGlyGlyHisGlyProLeu 839

Qy     850 ------GGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG------CTGGGC 897
                     |||||||||   |||||||||      :::|||      :::            |||
Db     840 ThrAspAspProGlyGlyAsnGlyGlyThrGlyGlyAsnGlyGlyThrGlyGlyThrGly 859

Qy     898 GGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTG--------------------------- 930
           |||   |||:::   :::|||   |||
Db     860 GlyAlaGlyIleGlySerLeuGlyGlyGlyThrGlyGlyAspGlyGlyAsnGlyGlyAsn 879
```

RESULT 10
S22697
extensin - Volvox carteri (fragment)
C;Species: Volvox carteri
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-Jul-2000
C;Accession: S22697; S21096
R;Ertl, H.; Hallmann, A.; Wenzl, S.; Sumper, M.
EMBO J. 11, 2055-2062, 1992
A;Title: A novel extensin that may organize extracellular matrix biogenesis in Volvox ca
A;Reference number: S22697; MUID:92289669; PMID:1600938
A;Accession: S22697
A;Molecule type: mRNA
A;Residues: 1-464 <HAL>
A;Cross-references: EMBL:X65165; NID:g21991; PIDN:CAA46283.1; PID:g21992
C;Keywords: glycoprotein

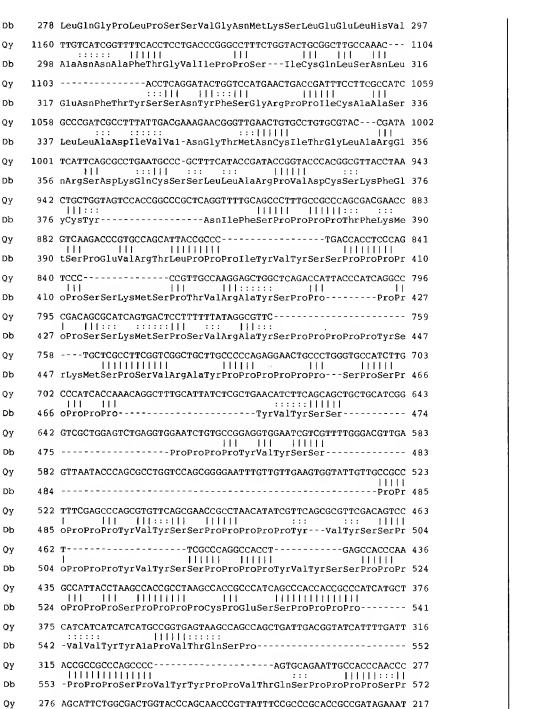
Alignment Scores:
Pred. No.: 2.78e-11 Length: 464
Score: 253.50 Matches: 116
Percent Similarity: 32.11% Conservative: 15
Best Local Similarity: 28.43% Mismatches: 99
Query Match: 9.78% Indels: 178
DB: 2 Gaps: 20

SEQ12-SEQ4 (1-1390) x S22697 (1-464)



RESULT 11
F70971
hypothetical glycine-rich protein Rv3367 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis

```
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 18-Feb-2000
C;Accession: F70971
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70971
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-588 <COL>
A;Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15752.1; PID:e120228
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3367
C;Superfamily: unassigned collagens

Alignment Scores:
Pred. No.:               2.87e-11      Length:         588
Score:                   253.50        Matches:        123
Percent Similarity:      36.18%        Conservative:   17
Best Local Similarity:   31.78%        Mismatches:     147
Query Match:             10.01%        Indels:         100
DB:                      2             Gaps:           21

SEQ12-SEQ4 (1-1390) x F70971 (1-588)

Qy    226 GGCGGTGCGGGCGGA------AATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGG 279
            ||||||||||||||        |||          ||||||     |||||||||
Db    283 GlyGlyAlaGlyGlyAspAlaAsnGlyAsnProAlaAsnThrSerIleAlaAsnAlaGly 302

Qy    280 TTGGGTGGCAATTCTGCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAG 339
               |||||||||  |||  |||  |||  |||     |||
Db    303 AlaGlyGlyAsnGlyAlaAlaGly---GlyAspGlyGlyAlaAsnGlyGlyAlaGlyGly 321

Qy    340 CTGGCTGGCTTACTCACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTG 399
                |||                      :::       ||||||  |||
Db    322 AlaGlyGlyGlnAlaAlaSerAlaGlySerSerValGlyGlyAspGlyGlyAsnGlyGly 341

Qy    400 ATGGGCGGT-----------GGCTTAGGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCA 447
             ||||||            |||    ||||||  |||  |||   ||||||:::
Db    342 AlaGlyGlyThrGlyThrAsnGlyHisAlaGlyGlyAlaGly---GlyAlaGlyGlyAla 360

Qy    448 GGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAAC 507
            ||||||  |||      |||            |||      ||||||:::
Db    361 GlyGlyArgGlyGlyTrpLeuValGly------AsnGlyGlyAsnGlyGlyAsn------ 376

Qy    508 ACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCCCCGCTGGACCAG 567
                |||:::  |||||||||              |||
Db    377 ------GlyAlaAlaGlyGlyAsnGlyAlaIleGlyGlyThrGlyGly----------- 390

Qy    568 GCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCCACC 627
            |||   |||:::       :::|||     :::|||     ||||||
Db    391 AlaGlyGlyVal---ProAlaAsnGlnGlyGlyAsnSerAlaLeuGlyThr--------- 406

Qy    628 TCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAAAGC 687
                                                                   |||
Db    407 -------------------------------------------------------GlnPro 408

Qy    688 CTGTTTGGTGATGGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACC 747
            :::  |||||||||  ||||||    |||  :::  |||            |||
Db    409 ValGlyGlyAspGlyGlyAspGlyGlyAsnGlyGlyThrGlyGly-----------Thr 424

Qy    748 GAAGGCGAGCAGAACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGC---CTGATG 804
               |||               |||           ||||||  ||||||
Db    425 GlyGlyArgGlyGlyAspGlyGlySerGlyGlyAlaGlyGlyAlaSerGlyTrpLeuMet 444

Qy    805 GGTAATGGTCTGAGCCAGCTCCTTGGCAACGGGGGACTGGGA------GGTGGTCAGGGC 858
            |||||||||   :::       ||||||||||||  |||      ||||||  |||
Db    445 GlyAsnGlyGlyAsnGly------GlyAsnGlyGlyThrGlyGlySerGlyGlyValGly 462

Qy    859 GGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTG 918
            ||||||   |||   |||    ||||||:::       ||||||
Db    463 GlyAsnGlyGlyIleGlyGlyAspGlyAlaGly---GlyGlyAsnAlaThrSerThrSer 481

Qy    919 AGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAA 978
            |||   |||   |||                 |||         |||   |||   |||  ::::
Db    482 SerIleProPheAspAlaHis------GlyGlyAsnGlyGlyAlaGly---Gly--AspA 498

Qy    979 GCGGGCATTCAGGCGCTGAATGATATCGGTACGCACAGGCACAGTTCAACCCGTTCTTTC 1038
            ::||||||   |||
Db    498 laGlyHisGlyGlyThr------------------------------------------ 503

Qy   1039 GTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATCGGTCAGTTCATGGACCAGTATCCT 1098
                                 |||||||||    |||:::            |||
Db    504 -------------GlyGlyAspGly---GlyAspGlyGlyHisAlaGlyThrGlyGlyA 518

Qy   1099 GAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCGGGTCAGGAGGTGAAAACCGATGAC 1158
               |||         |||                |||:::||||||
Db    518 rgGlyGlyLeuLeuAlaGlyGlnHisAlaAsnSerGlyAsnGlyGly------------ 533

Qy   1159 AAATCATGGGCAAAAGCACTGAGCAAGCCAGATGACGACGGAATGACACCAGCCAGTATG 1218
                   |||       |||       |||                ||||||    |||
Db    534 --------GlyGlyGlyThrGlyGlyAlaGlyGlyThrHisGly-ThrProGlySerGly 550

Qy   1219 GAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGGCCCATGGCGGGTGATACCGGCAAC 1278
                                                        ||||||    ||||||
Db    551 Asn-------------------------------------AlaGlyGlyThrGlyThr 557

Qy   1279 GGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCGCTGGGTATTGATGCCATGATGGCC 1338
            ||||||      :::          |||       |||    |||              :::
Db    558 GlyAsnAlaAspSerThrAsnGlyGlyProGlySerAspGly-----------LeuGly 573

Qy   1339 GGTGATGCCATTAAC 1353
            |||||||||    |||
Db    574 GlyAspAlaPheAsn 578

RESULT 12
E86255
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: E86255
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E86255
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-744 <STO>
A;Cross-references: GB:AE005172; NID:g3157926; PIDN:AAC17609.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Alignment Scores:
Pred. No.:               2.96e-11      Length:         744
Score:                   253.50        Matches:        110
Percent Similarity:      38.44%        Conservative:   43
Best Local Similarity:   27.64%        Mismatches:     119
Query Match:             9.78%         Indels:         127
DB:                      2             Gaps:           21

SEQ12-SEQ4 (1-1390) x E86255 (1-744)

Qy   1214 CTGGCTGGTGTCATTCCGTCGTCATCTGGCTTGCTCAGTGCTTTT------GCCCATGAT 1161
            |||   |||   :::|||||||||   |||   :::   :::            |||
```



Db 278 LeuGlnGlyProLeuProSerSerValGlyAsnMetLysSerLeuGluGluLeuHisVal 297

Qy 1160 TTGTCATCGGTTTTCACCTCCTGACCCGGGCCTTTCTGGTACTGCGGCTTGCCAAAC--- 1104

Db 298 AlaAsnAsnAlaPheThrGlyValIleProProSer---IleCysGlnLeuSerAsnLeu 316

Qy 1103 ----------ACCTCAGGATACTGGTCCATGAACTGACCGATTTCCTTCGCCATC 1059

Db 317 GluAsnPheThrTyrSerSerAsnTyrPheSerGlyArgProProIleCysAlaAlaSer 336

Qy 1058 GCCCGATCGCCTTTATTGACGAAAGAACGGCTTGAACTGTGCCTGTGCGTAC---CCATA 1002

Db 337 LeuLeuAlaAspIleValVal-AsnGlyThrMetAsnCysIleThrGlyLeuAlaArgGl 356

Qy 1001 TCATTCAGCGCCTGAATGCCC-GCTTTCATACCGATACCGGTACCCACGGCGTTACCTAA 943

Db 356 nArgSerAspLysGlnCysSerSerLeuLeuAlaArgProValAspCysSerLysPheGl 376

Qy 942 CTGCTGGTAGTCCCACGGCCCGCTCAGGTTTTGCAGCCCTTTGCCGCCCAGCGACGAACC 883

Db 376 yCysTyr----------AsnIlePheSerProProProThrPheLysMe 390

Qy 882 GTCAAGACCCGTGCCAGCATTACCGCCC----------TGACCACCTCCCAG 841

Db 390 tSerProGluValArgThrLeuProProProIleTyrValTyrSerSerProProProPr 410

Qy 840 TCCC----------CCGTTGCCAAGGAGCTGGCTCAGACCATTACCCATCAGGCC 796

Db 410 oProSerSerLysMetSerProProThrValArgAlaTyrSerProPro---------ProPr 427

Qy 795 CGACAGCGCATCAGTGACTCCTTTTTTATAGGCGTTC---------- 759

Db 427 oProSerSerLysMetSerProSerValArgAlaTyrSerProProProProProTyrSe 447

Qy 758 ---TGCTCGCCTTCGGTCGGCTGCTTGCCCCCAGAGGAACTGCCCTGGGTGCCATCTTG 703

Db 448 rLysMetSerProSerValArgAlaTyrProProProProProPro---SerProSerPr 466

Qy 702 CCCATCACCAAACAGGCTTTGCATTATCTCGCTGAACATCTTCAGCAGCTGCTGCATCGG 643

Db 466 oProProPro----------TyrValTyrSerSer---------- 474

Qy 642 GTCGCTGGAGTCTGAGGTGGAATCTGTGCCGGAGGTGGAATCGTCGTTTTGGGACGTTGA 583

Db 475 ----------ProProProProTyrValTyrSerSer---------- 483

Qy 582 GTTAATACCCAGCGCCTGGTCCAGCGGGAATTTGTTGTTGAAGTGGTATTGTTGCCGCC 523

Db 484 ----------ProPr 485

Qy 522 TTTCGAGCCCAGCGTGTTCAGCGAACCGCCTAACAATATCGTTCAGCGCGTTCGACAGTCC 463

Db 485 oProProProTyrValTyrSerSerProProProProProTyr---ValTyrSerSerPr 504

Qy 462 T----------TCGCCCAGGCCACCT----------GAGCCACCCAA 436

Db 504 oProProProTyrValTyrSerSerProProProProTyrValTyrSerSerProProPr 524

Qy 435 GCCATTACCTAAGCCACCGCCTAAGCCACCGCCCATCAGCCCACCACCGCCCATCATGCT 376

Db 524 oProProProSerProProProProProCysProGluSerSerProProProProPro-------- 541

Qy 375 CATCATCATCATCATGCCGGTGAGTAAGCCAGCCAGCTGATTGACGGTATCATTTTGATT 316

Db 542 -ValValTyrTyrAlaProValThrGlnSerPro---------- 552

Qy 315 ACCGCCGCCCAGCCCC----------AGTGCAGAATTGCCACCCAACCC 277

Db 553 -ProProProSerProValTyrTyrProProValThrGlnSerProProProProSerPr 572

Qy 276 AGCATTCTGGCGACTGGTACCCAGCAACCCGTTATTTCCGCCGCACGCCGATAGAAAT 217

Db 572 oValTyrTyrProProValThrAsnSerPro---------- 584

Qy 216 TTGCATCGTTGACGCTCCCAGCCCACTTGTATTCAGACTCATAACGTATTCC 165

Db 585 ----------ProSerProValTyrTyrProProValThrTyrSer 596

RESULT 13
C29356
hydroxyproline-rich glycoprotein (clone Hyp2.13) - kidney bean (fragment)
C;Species: Phaseolus vulgaris (kidney bean)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 21-Jul-2000
C;Accession: C29356
R;Corbin, D.R.; Sauer, N.; Lamb, C.J.
Mol. Cell. Biol. 7, 4337-4344, 1987
A;Title: Differential regulation of a hydroxyproline-rich glycoprotein gene family in
A;Reference number: A29356; MUID:88142825; PMID:3437892
A;Accession: C29356
A;Molecule type: mRNA
A;Residues: 1-368 <COR>
A;Cross-references: EMBL:M18095; NID:g169348; PIDN:AAA33765.1; PID:g169349
A;Experimental source: cv. Kievitsboon Koekoek
C;Superfamily: hydroxyproline-rich glycoprotein
C;Keywords: glycoprotein

Alignment Scores:

Pred. No.:	4.17e-11	Length:	368
Score:	251.00	Matches:	108
Percent Similarity:	31.04%	Conservative:	23
Best Local Similarity:	25.59%	Mismatches:	97
Query Match:	9.68%	Indels:	195
DB:	2	Gaps:	23

SEQ12-SEQ4 (1-1390) x C29356 (1-368)

Qy 1387 CTTAAGCCGCCGCCCAGCTTGCCAAGTGCCATATTGTTAATGGCATCACCGGCCATCATGG 1328

Db 81 LeuLysAlaLysProPheAlaTyrAlaPro---------------LysLysPro----- 93

Qy 1327 CATCAATACCCAGCGAAGAACCACCGGCACCGCGTGCCTGCAGGTTGCCGTTGCCGGTAT 1268

Db 94 ------TyrAspCysGluLysHisLysHisSerProThrPro---------------Tyr 106

Qy 1267 CACCCGCCATGGGCCTTTTGATCATGCCCTTGGCTTTGTTGAACTGCTCCATACTGGCTG 1208

Db 107 HisLysProTyrTyrTyrAsnSerProPro---------------ProProTyrTyrTyr 121

Qy 1207 GTGTCATTCCGTCGTCATCTGGCTTGCTCAGTGCTTTTGCCCATGATTTGTCATCGGTTT 1148

Db 122 LysSer----------ProProTyrTyrTyrLys---- 129

Qy 1147 TCACCTCCTGACCCGGGCCTTTCTGGTACTGCGGCTTGCCAAACACCTCAGGATACTGGT 1088

Db 130 SerProProProProSer---------- 135

Qy 1087 CCATGAACTGACCGATTTCCTTCGCCATCGCCCGATCGCCTTTATTGACGAAAGAACGGG 1028

Db 136 ----------ProSerProSerPro---------- 140

Qy 1027 TTGAACTGTGCCTGTGCGTACCGATATCATTCAGCGCCTGAATGCCCGCTTTCATACCGA 968

Db 141 ----------TyrTyrTyrLysSerProProProProProHisLysAspPro 153

Qy 967 TACCGGTACCCACGGCCGTTACCTAACTGCTGGTAGTCCACCGGC---------- 924

Db 154 Tyr---TyrProProTyrTyrTyrLysSerProProProProSerProSerProSerProPro 172

Qy 923 CCGCTCAGGTTTTGCAGCCCTTTGCCGCCCAGCGACGAACCGTCAAGACCCGTGCCA--- 867

Db 173 ProTyrTyrTyrLysSerProProProProProSer------ProSerProProProTyr 190

Qy 866 ------GCATTACCGCCCTGACCACCTCCCAGTCCCCCGTTGCCAAGGAGCTGGCTCAGA 813

Db 191 TyrTyrLysSerProProProProAspProSerProProProProProTyrTyrTyrLysSer 210

```
Qy    812 CCATTACCCATCAGGCCCGACAGCGCATCAGTGACTCCTTTTTTATAGGCGTTCTGCTCG 753
          |||  |||     |||                                       |||
Db    211 ProProProProSerPro-------------------------------------Ser 217

Qy    752 CCTTCGGTCGGCTGCTTGCCCCCAGAGGAACTGCCCTGGGTGCCATCTTGCCCATCACCA 693
          |||              ||||||            |||     |||||||||
Db    218 Pro--------------ProProProTyrTyrTyrLysSerProProProProSerPro 232

Qy    692 AACAGGCTTTGCATTATCTCGCTGAACATCTTCAGCAGCTGCTGCATCGGGTCGCTGGAG 633
          :::
Db    233 Ser--------------------------------------------------------- 233

Qy    632 TCTGAGGTGGAATCTGTGCCGGAGGTGGAATCGTCGTTTGGGACGTTGAGTTAATACCC 573

Db    233 ------------------------------------------------------------ 233

Qy    572 AGCGCCTGGTCCAGCGGGGAATTTGTTGTTGAAGTGGTATTGTTGCCGCCTTTCGAGCCC 513
                                                  ||||||    |||
Db    234 ------------------------------------------ProPro---ProPro 237

Qy    512 AGCGTGTTCAGCGAACCGCCTAACATATCGTTCAGCGCGTTCGACAGTCCTTCGCCCAGG 453
             :::      ||||||   |||   |||        |||||||||||
Db    238 TyrTyrTyrLysSerProProProProSerProSerProProProSerProSerProPro 257

Qy    452 CCA-----------------CCTGAGCCACCCAAGCCATTACCTAAGCCACCG------ 417
          |||                 |||   ||||||   |||    |||    ||||||
Db    258 ProProTyrTyrTyrLysSerProProProProAspProSerProProProProTyrTyr 277

Qy    416 --------CCTAAGCCACCGCCCATCAGCCCACCACCGCCCATCATGCTCATCATCATC 366
                  |||   ||||||      ||||||||||||||
Db    278 TyrLysSerProProProProSerProSerProProProProTyrTyr----------- 293

Qy    365 ATCATGCCGGTGAGTAAGCCAGCCAGCTGATTGACGGTATCATTTTGATTACCGCCGCCC 306
                          |||                            |||||||||
Db    294 ---------TyrLysSerPro----------------------------ProProPro 300

Qy    305 AGCCCCAGTGCAGAATTGCCACCCAACCCAGCATTCTGGCGACTGGTACCCAGCAACCCG 246
          |||||||||        ||||||  |||   ::::::::      |||
Db    301 SerProSer---------ProProProProTyrTyrTyrLysSerProPro--------- 314

Qy    245 TTATTTCCGCCCGCACCGCCGATAGAAATTTGCATCGTTGACGCTCCCAGCCCACTTGTA 186
                ||||||:::|||                     |||    |||:::|||
Db    315 ------ProProSerPro-------------------AspProProThrProTyrTyr 325

Qy    185 TTCAGA 180
          ::::::
Db    326 TyrLys 327
```

RESULT 14
C70974
hypothetical glycine-rich protein Rv3388 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: C70974
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: C70974
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-731 <COL>
A;Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15773.1; PID:g266166
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3388
C;Superfamily: elastin

```
Alignment Scores:
Pred. No.:              6.47e-11      Length:        731
Score:                  249.00        Matches:        96
Percent Similarity:     35.63%        Conservative:   23
Best Local Similarity:  28.74%        Mismatches:    101
Query Match:            9.83%         Indels:        114
DB:                     2             Gaps:           15
```

SEQ12-SEQ4 (1-1390) x C70974 (1-731)

```
Qy    178 AGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGC 237
          |||      |||    |||:::|||||||||                  ||||||   |||
Db    402 SerAlaGlyThrGlyGlyValGlyAlaSer------------------GlyGlyThrGly 415

Qy    238 GGAAATAACGGGTTGCTGGGTACCAGTCGCCAG--------------------------- 270
          ||||||   ||||||:::|||
Db    416 GlyAsnAlaGlyLeuIleGlyAlaGlyGlyHisGlyGlyAlaGlyGlyAlaGlyGlyAsn 435

Qy    271 ------------------AATGCTGGGTTGGGTGGCAAT---TCTGCACTGGGGCTGGGC 309
                            |||   |||   |||||||||      |||   |||   |||
Db    436 GlnThrGlyGlyValGlyAsnGlyGlyAlaGlyGlyAsnGlyGlyAlaGlyGlyAlaGly 455

Qy    310 GGC---------GGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGC 360
          |||         ||||||       |||        |||       ||||||
Db    456 GlyGlnLeuTyrGlyAsnGlyGlyAspGlyGlyAsnGlyGlyAlaGly------------ 471

Qy    361 ATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGC--- 417
                   ::::::   ||||||   |||      ||||||      |||
Db    472 ---------GlyAlaAsnIleAlaGlyGlyAsnGlySerAspGlyGlyAlaAlaGlyHis 488

Qy    418 GGTGGCTTAGGT-----------------------AATGGCTTGGGTGGCTCAGGTGGC 453
          ||||||   |||                          |||   ||||||   ||||||
Db    489 GlyGlyAlaGlyGlySerAlaArgLeuIleGlyAlaGlyGlyHisGlyGlyAspGlyGly 508

Qy    454 CTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTG 513
          |||          :::         |||   :::   |||:::
Db    509 AlaGlyGlyAsnThrAlaGlyArgArgAlaAspAlaIleAlaGlyThr------------ 524

Qy    514 GGCTCGAAAGGCGGCAAC------------------------------------------ 531
          |||    |||||||||
Db    525 GlyGlyAspGlyGlyAsnGlyGlyAsnGlyGlyLeuLeuSerGlyAsnAlaGlyAlaGly 544

Qy    532 --------------------------AATACCACTTCAACAACAAATTCCCCGCTGGAC 564
                                       |||:::||||||   :::|||
Db    545 GlyHisGlyGlyAlaGlyGlySerSerThrAlaThrThrThrThrGlyThrProProThr 564

Qy    565 CAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGCACAGATTCC 624
             |||   |||   |||         :::                  |||:::|||
Db    565 GlyAlaThrGlyGlyAsnGlyGlyAsnGlyGlyAlaGlyGlyThrAlaGlyPheThrGly 584

Qy    625 ACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAGATAATGCAA 684
          :::
Db    585 SerGly------------------------------------------------------ 586

Qy    685 AGCCTGTTTGGTGATGGGCAAGATGGCACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCG 744
             :::   |||:::|||      |||         |||    ::::::|||
Db    587 GlyIleGlyGlyAsnGlyGlyAlaGlyGlyThrGlyGlyAsnAlaGly------------ 602

Qy    745 ACCGAAGGCGAGCAGAACGCCTATAAAAAAGGAGTCACTGATGCGCTG------------ 792
                                   |||       |||    |||       |||
Db    603 ---------------ValAlaLeuSerValGlySerThrGlyGlyLeuGlyGlyAsnGly 617

Qy    793 ---TCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAACGGGGGACTGGGA--- 846
             ||||||   :::|||    |||          |||    |||||||||||   |||
Db    618 GlySerGlyGlyLeuGlyGlyGlyGlyGlySerLeuPheGlyAsnGlyGlyAlaGlyGly 637

Qy    847 ---GGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTGGGCGGCAAA 903
             |||      |||||||||   |||:::|||:::         :::|||:::||||||
Db    638 ValGlyAlaThrGlyGlyAsnGlyGlySerGlyIleGlyProAlaSerValGlyGlyAsn 657

Qy    904 GGG---------------------CTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAG 942
          |||                        |||:::|||                   |||
Db    658 GlyGlyLysGlyGlyValGlyAlaAlaGlyGlyLeuAlaGly---------------Gln 672

Qy    943 TTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCGGGC 984
```

```
              :::||||||           |||    |||         |||
Db      673 IleGlyAsnGlyGlySerGlyGlySerGlyGlyAlaGlyGly 686

RESULT 15
B70893
hypothetical glycine-rich protein Rv1068c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: B70893
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: B70893
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-463 <COL>
A;Cross-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA17184.1; PID:g28967C
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1068c
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Alignment Scores:
Pred. No.:               7.25e-11     Length:          463
Score:                   248.00       Matches:         93
Percent Similarity:      37.67%       Conservative:    17
Best Local Similarity:   31.85%       Mismatches:      112
Query Match:             9.79%        Indels:          70
DB:                      2            Gaps:            13

SEQ12-SEQ4 (1-1390) x B70893 (1-463)

Qy      196 CTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCGGGCGGAAATAACGGGTTGCTG 255
              :::||||||     :::           ||||||||||||||       ||||||
Db      203 IleGlyAlaProGlyValAlaGlyGlyAlaGlyGlyAlaGlyGlyThrAlaGlyLeuPhe 222

Qy      256 GGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCTGCACTGGGGCTGGGCGGCGGT 315
             |||                 |||   |||   ||||||             |||      ||||||
Db      223 Gly------------AsnGlyGlyAlaGlyGlyAlaGlyGlyAlaGlyGlyAlaGlyGly 238

Qy      316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                         |||               ||||||    |||:::|||
Db      239 ArgGlyGlyAsp------GlyGlySerAlaGlyTrpLeuSerGlyAsnGlyGlyAspAla 256

Qy      376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                         ||||||||||||       |||    |||       ||||||
Db      257 GlyThrGlyGlyGlyGlyGlyAsnAlaGlyAsnGlyGlyAsnGlyGlySerAlaGlyTrp 276

Qy      436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
             |||    |||:::||||||    |||   |||                           |||
Db      277 LeuSerGlyAsnGlyGlyThrGlyGlyGly---------------------------Gly 287

Qy      496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
             |||:::               |||   :::|||||||||                ||||||
Db      288 GlyThrAlaGlyAla---GlyGlyGlnGlyGlyAsnGly---------------AsnSer 301

Qy      556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                 :::|||                |||                |||       :::|||
Db      302 GlyIleAspProGly---------AsnGlyGlyGlnGlyAlaAspThrGlyAsnAlaGly 318

Qy      616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
                                |||:::
Db      319 AsnGlyGlyHisGlyGlySerAla------------------------------------ 326

Qy      676 ATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACCCAGGGCAGTTCC--------- 726
                          |||||||||||||||        |||      |||
Db      327 ------AlaLysLeuPheGlyAspGlyGlyAlaGlyGlyAlaGlyGlyMetGlySerThr 344

Qy      727 --------------------------TCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAG 759
                                                ||||||              |||
Db      345 GlyGlyThrGlyGlyGlyGlyGlyPheGlyGlyGlyThrGlyGlyAsnGlyGlyAsnGly 364

Qy      760 AACGCCTATAAAAAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGC 819
              :::|||         |||   :::          :::||||||:::|||:::|||   |||
Db      365 HisAlaGlyGlyAlaGlyGlySerGlyGlyThrAlaGlyLeuLeuGlySerGlyGlySer 384

Qy      820 CAGCTCCTT------------GGCAACGGGGGACTGGGAGGTGGTCAGGGC------GGT 861
                                  ||||||||||||||||||    |||    |||       |||
Db      385 GlyGlyThrGlyGlyAspGlyGlyAsnGlyGlyLeuGlyAlaGlySerGlyAlaLysGly 404

Qy      862 AATGCTGGCACGGGTCTTGACGGTTCGTCGCTGGGCGGCAAAGGGCTGCAAAACCTGAGC 921
             |||    |||    |||    |||                 |||||||||||
Db      405 AsnGlyGlyAsnGlyGlyAsp------------GlyGlyLysGly--------------- 415

Qy      922 GGGCCGGTGGACTACCAGCAGTTAGGTAACGCCGTGGGTACCGGTATCGGTATGAAAGCG 981
                         |||    |||    :::||||||           |||    |||    |||
Db      416 ------GlyAspAlaGlnLeuIleGlyAsnGlyGlyAsnGlyGlyAsnGlyGlyLysGly 433

Qy      982 GGC---------ATTCAGGCGCTGAATGATATCGGT 1008
             |||          :::        :::|||      |||
Db      434 GlyThrGlyLeuMetProGlyIleAsnGlyThrGly 445

Search completed: January 14, 2003, 17:12:24
Job time : 41 secs
```

102 n.a. 1288 n.a.

GenCore version 5.1.3

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 14, 2003, 16:59:17 ; Search time 40.5 Seconds
(without alignments)
9146.586 Million cell updates/sec

Title: SEQ12-SEQ4
Perfect score: 2532
Sequence: 1 atgggatttttctctttttc.....gctgggcgcggcttaagctt 1390

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/KUBELIK693/runat_14012003_091818_7057/app_query.fasta_1.1543
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=KUBELIK693_@CGN_1_1_51_@runat_14012003_091818_7057 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2079	82.1	403	19	AAW75863	Erwinia amylovora
	2	2079	82.1	403	19	AAW62455	Erwinia amylovora
	3	2079	82.1	403	19	AAW61114	Hypersensitive res
	4	2079	82.1	403	20	AAW87639	A hypersensitive r
	5	2079	82.1	403	21	AAY71093	Erwinia amylovora
	6	2079	82.1	403	21	AAY84854	A hypersensitive r
	7	2079	82.1	403	22	AAE06710	Erwinia amylovora
	8	2079	82.1	403	23	ABB09225	Erwinia amylovora
	9	2079	82.1	403	23	AAE18295	Erwinia amylovora
	10	2079	82.1	403	23	AAE16447	E. amylovora hyper
	11	1986	78.4	385	18	AAW06598	Hypersensitive res
	12	1962	77.5	385	15	AAR45751	Erwinia amylovora
	13	718.5	28.4	338	18	AAW06597	Hypersensitive res
	14	718.5	28.4	338	20	AAW87638	A hypersensitive r
	15	718.5	28.4	338	20	AAW82407	E. chrysanthemi Hr
	16	718.5	28.4	338	21	AAY71092	Erwinia chrysanthe
	17	718.5	28.4	338	21	AAY84853	A hypersensitive r
	18	718.5	28.4	338	21	AAY55801	E. chrysanthemi hy
	19	718.5	28.4	338	22	AAR06709	Erwinia chrysanthe
	20	718.5	28.4	338	23	ABB09224	Erwinia chrysanthe
	21	718.5	28.4	338	23	AAE18294	Erwinia chrysanthe
	22	718.5	28.4	338	23	AAE16446	E. chrysanthemi hy
	23	718.5	28.4	340	19	AAW75862	Erwinia chrysanthe
	24	718.5	28.4	340	19	AAW61113	Hypersensitive res
	25	713.5	28.2	340	19	AAW62454	Erwinia chrysanthe
c	26	298	11.5	572	18	AAW31855	Mycobacterium tube
c	27	298	11.5	763	18	AAW31852	Mycobacterium tube
c	28	281.5	10.9	406	22	ABG27250	Novel human diagno
c	29	253.5	9.8	598	22	ABG14000	Novel human diagno
c	30	250.5	9.7	325	22	ABG21919	Novel human diagno
c	31	235.5	9.1	694	23	ABB92421	Herbicidally activ
c	32	234.5	9.0	783	19	AAW37151	Mouse neural Mena+
c	33	234.5	9.0	787	19	AAW37152	Mouse neural Mena+
c	34	234.5	9.0	802	19	AAW37153	Mouse neural Mena+
c	35	234.5	9.0	802	22	AAU09139	Mammalian enabled
	36	226.5	8.9	898	18	AAW31853	Mycobacterium tube
	37	222	8.8	272	22	ABB70198	Drosophila melanog
c	38	220	8.5	731	22	AAB74209	Protein encoded by
c	39	220	8.5	731	23	ABB93202	Herbicidally activ
	40	217.5	8.6	718	12	AAR14308	N.clavipes draglin
c	41	216	8.3	439	13	AAR28150	Sugar beet chitina
	42	213.5	8.4	604	16	AAR99057	Spider dragline va
	43	213.5	8.4	651	20	AAY14097	Spider silk protei
	44	213.5	8.4	651	23	AAU11781	Spider natural sil
	45	213.5	8.4	718	19	AAW53346	Nephila clavipes s

ALIGNMENTS

RESULT 1
AAW75863
ID AAW75863 standard; Protein; 403 AA.
XX
AC AAW75863;
XX
DT 07-DEC-1998 (first entry)
XX
DE Erwinia amylovora hypersensitive response elicitor (HRE).
XX
KW Hypersensitive response elicitor; HRE; insect resistance;
KW biological control; transgenic plant.
XX
OS Erwinia amylovora.
XX
PN WO9837752-A1.
XX
PD 03-SEP-1998.
XX

```
PF   26-FEB-1998;   98WO-US03604.
XX
PR   28-FEB-1997;   97US-0039226.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Wei Z,  Zitter TA;
XX
DR   WPI; 1998-495374/42.
DR   N-PSDB; AAV54607.
XX
PT   Use of hypersensitive response elicitor polypeptide - for
PT   application to plants or seeds or transgenic plants or seeds for the
PT   control of insects.
XX
PS   Disclosure; Page 9-10; 75pp; English.
XX
CC   This is the amino acid sequence of a 39 kDa, heat stable
CC   hypersensitive response elicitor (HRE) of Erwinia amylovora. The
CC   invention relates to the use of a HRE polypeptide or protein to
CC   control insects on plants or plants grown from seed treated with HRE.
CC   Also claimed is a method of insect control for plants that involves:
CC   (a) providing a transgenic plant or seed transformed with a DNA
CC   molecule (see AAV54606-09) encoding a HRE polypeptide or protein (see
CC   AAW75862-67); and (b) growing the transgenic plants or transgenic
CC   plants produced from the transgenic seeds to control insects.  HRE
CC   prevents direct insect damage to plants by feeding injury.  It kills
CC   insects close to plants, and interferes with insect larval feeding
CC   on such plants.  It also prevents insects from colonising host
CC   plants and releasing phytotoxins which result in disease damage to
CC   plants.
XX
SQ   Sequence   403 AA;

Alignment Scores:
Pred. No.::          1.61e-161       Length:          403
Score:               2079.00         Matches:         403
Percent Similarity:  100.00%         Conservative:    0
Best Local Similarity: 100.00%       Mismatches:      0
Query Match:         82.11%          Indels:          0
DB:                  19              Gaps:            0

SEQ12-SEQ4 (1-1390) x AAW75863 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy    235 GGCGGAAATAACGGGTTGCTGGCTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy    295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy    355 ACCGGCATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGCACTCCAGCGACCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCACGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGCGGTAATGCTGGCACGGGTCTTGCAGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGGTACCGGTATCGGTATGAAAGCGGGCAATTCAGGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATCGCCGCTGATGCCATTAACAATATGGCACTTGGCAACGTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db    401 GlyAlaAla 403

RESULT 2
AAW62455
ID   AAW62455 standard; Protein; 403 AA.
XX
AC   AAW62455;
XX
DT   09-NOV-1998  (first entry)
XX
DE   Erwinia amylovora hypersensitive response elicitor (HRE).
XX
KW   Hypersensitive response elicitor; HRE; growth; transgenic plant.
XX
OS   Erwinia amylovora.
XX
PN   WO9832844-A1.
XX
PD   30-JUL-1998.
XX
```

```
PF   27-JAN-1998;   98WO-US01507.
XX
PR   27-JAN-1997;   97US-0036048.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PT   Beer SV,  Qiu D,  Wei Z;
XX
DR   WPI; 1998-427940/36.
DR   N-PSDB; AAV39973.
XX
PT   Method for enhancing plant growth - comprises use of hypersensitive
PT   response elicitor polypeptide or protein which may also effect, e.g.
PT   increase in plant height or earlier germination seed
XX
PS   Disclosure; Page 15-16; 110pp; English.
XX
CC   This is the deduced amino acid sequence of the 39 kDa, heat stable
CC   hypersensitive response elicitor (HRE) of Erwinia amylovora.  A
CC   method of enhancing growth in plants comprises: (a) applying a HRE
CC   polypeptide or protein in a non-infectious form to a plant or plant
CC   seed under conditions effective to enhance growth of the plant or
CC   plants grown from the seed, or (b) providing a transgenic plant or
CC   plant seed transformed with a DNA molecule encoding a HRE
CC   polypeptide or protein, and growing the transgenic plant or a plant
CC   produced from the transgenic seed under conditions effective to
CC   enhance plant growth.  HREs (see AAW62454-59) or nucleic acids
CC   encoding them (see AAV39972-75) can be used to increase plant growth.
CC   The HREs may also result in increased plant height and yield, and
CC   effect early germination and maturation of plant seed and early
CC   colouration of fruit and plants.  E. amylovora HRE can be applied
CC   to tomato plants to enhance growth without causing disease in that
CC   species; this bacterium is a pathogen of apple and pear but not
CC   of tomato.
XX
SQ   Sequence   403 AA;

Alignment Scores:
Pred. No.:              1.61e-161       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
DB:                     19              Gaps:           0

SEQ12-SEQ4 (1-1390) x AAW62455 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyTrpGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy    235 GGCGGAAATAACGGCTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTCGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy    295 GCACTGGGGCTCGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy    355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlyGlyLeuGlyGlyGlyLeuGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGACCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGGTACCGGTATCGGTATGAAAGCGGCCATTCAGGCGCTGAATGATATCGGTACCCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGCAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTGAAAAACGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db    401 GlyAlaAla 403  .

RESULT 3
AAW61114
ID   AAW61114 standard; Protein; 403 AA.
XX
AC   AAW61114;
XX
DT   26-OCT-1998  (first entry)
XX
DE   Hypersensitive response elicitor protein (39 kDa).
XX
KW   Hypersensitive response elicitor; transgenic plant; seed;
KW   pathogen resistance; disease resistance; crop protection.
XX
OS   Erwinia amylovora.
XX
```

```
PN   WO9824297-A1.
XX
PD   11-JUN-1998.
XX
PF   04-DEC-1997;   97WO-US22629.
XX
PR   05-DEC-1996;   96US-0033230.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Beer SV,  Qiu D,  Wei Z;
XX
DR   WPI; 1998-332931/29.
DR   N-PSDB; AAV36428.
XX
PT   Imparting pathogen resistance to plants - by applying a
PT   hypersensitive response elicitor polypeptide to seeds
XX
PS   Disclosure; Page 18-20; 85pp; English.
XX
CC   This is the 39 kDa hypersensitive elicitor (HRE) protein of
CC   Erwinia amylovora.  It is heat stable at 100 degC for at least 10
CC   min, has a pI of approximately 4.3, and contains substantially no
CC   cysteine.  The invention relates to methods of imparting
CC   hypersensitive response induced resistance to plants by treatment
CC   of seeds.  Isolated HRE proteins can be applied to seeds as a means
CC   of imparting pathogen resistance to plants grown from the seeds.
CC   Alternatively, bacteria containing the gene encoding the HRE can be
CC   applied to the plant seeds, or transgenic plant seeds containing a
CC   DNA molecule encoding an HRE polypeptide or protein are used.  HRE
CC   polypeptide sequences from Erwinia chrysanthemi, Erwinia amylovora,
CC   Pseudomonas syringae, Pseudomonas solanacearum, Xanthomonas
CC   campestris pv. glycines and Xanthomonas campestris pelargonii (see
CC   AAW61113-18) are provided.  The methods can impart pathogen
CC   resistance without using agents which are harmful to the
CC   environment or pathogenic to the plant seed being treated, or to
CC   adjacent plants.  E. amylovora causes disease in apple or pear
CC   but not tomato.  However, it elicits a hypersensitive response in
CC   tomato.  Thus, E. amylovora can be applied to tomato seeds to
CC   impart pathogen resistance without causing diseases in plants of
CC   that species.
XX
SQ   Sequence   403 AA;

Alignment Scores:
Pred. No.:              1.61e-161       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
DB:                     19              Gaps:           0

SEQ12-SEQ4 (1-1390) x AAW61114 (1-403)

Qy     175 ATGAGTCTGAATACAAGTGGGCTGGGACGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla  20

Qy     235 GGCGGAAATAACGGGTTGCTGGCTACCAGTCGCCAGAATGCTGGCTTGGGTGGCAATTCT 294
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer  40

Qy     295 GCACTGGGGCTGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu  60

Qy     355 ACCGGCATGATGATGATGACCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu  80

Qy     415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy     475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy     535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy     595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy     655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy     715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy     775 GGAGTCACTGATGGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy     835 GGGGGACTGGGAGTGGTCAGGGCGGTAATGCTGGCACGGTCTTGACGGTTCGTCGCTG 894
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GlyGlyLeuGlyGlyGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeuGlyAsn 240

Qy     895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy     955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy     1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCCAAGGAAATC 1074
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy     1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCCGCAGTACCAGAAAGGCCG 1134
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy     1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy     1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy     1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy     1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy     1375 GGCGCGGCT 1383
            |||||||||
Db     401 GlyAlaAla 403

RESULT 4
AAW87639
ID   AAW87639 standard; Protein; 403 AA.
XX
AC   AAW87639;
XX
DT   09-MAR-1999  (first entry)
XX
```

```
DE   A hypersensitive response elicitor protein.
XX
KW   Hypersensitive response elicitor protein; hairpin protein;
KW   disease resistance; seed quality; insect control; corn borer;
KW   Lepidoptera larvae; transgenic plant.
XX
OS   Erwinia amylovora.
XX
PN   WO9854214-A2.
XX
PD   03-DEC-1998.
XX
PF   28-MAY-1998;   98WO-US10874.
XX
PR   30-MAY-1997;   97US-0048109.
XX
PA   (CORR ) CORNELL RES FOUND INC.
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Beer SV,  Laby RJ,  Wei,Z;
XX
DR   WPI; 1999-070210/06.
DR   N-PSDB; AAV83989.
XX
PT   New fragments of an Erwinia hypersensitive response elicitor protein
PT   and related DNA - used to impart disease resistance to plants, to
PT   increase their growth and to control insects
XX
PS   Claim 4; Page 10-11; 94pp; English.
XX
CC   The present sequence represents a hypersensitive response elicitor
CC   protein (also called hairpin protein) that is able to elicit a
CC   hypersensitive response in plants. The specification also describes
CC   hypersensitive response elicitors from other pathogenic organisms.
CC   The protein, in non-infectious form, is applied to plants to impart
CC   disease resistance (to a wide range of viral, bacterial and fungal
CC   pathogens), to improve growth (yield, quantity and quality of seeds,
CC   to provide earlier germination etc.) and to control insects (e.g. corn
CC   borers, Lepidoptera larvae etc.) The same results are provided by
CC   transgenic plants expressing the protein.
XX
SQ   Sequence   403 AA;

Alignment Scores:
Pred. No.:              1.61e-161       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
DB:                     20              Gaps:           0

SEQ12-SEQ4 (1-1390) x AAW87639 (1-403)

Qy     175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy     235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy     295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTTACTC 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy     355 ACCGGCATGATGATGATGATGGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy     415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy     475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy     535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy     595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy     655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy     715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy     775 GGAGTCACTGATGCGCTGTCGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy     835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy     895 GGCGGCAAAGGGCTGCAAAACCTGACGCGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy     955 GTGGGTACCGGTATCGGTATGAAAGCGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy    1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy    1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy    1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAGCACTGAGCAAGCCAGATGAC 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy    1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy    1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy    1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy    1375 GGCGCGGCT 1383
           |||||||||
Db     401 GlyAlaAla 403

RESULT 5
AAY71093
ID   AAY71093 standard; Protein; 403 AA.
XX
AC   AAY71093;
XX
DT   08-SEP-2000  (first entry)
XX
DE   Erwinia amylovora hypersensitive response elicitor #1.
```

```
XX
KW   Hypersensitive response elicitor; environmental stress resistance;
KW   plant.
XX
OS   Erwinia amylovora.
XX
PN   WO2000028055-A2.
XX
PD   18-MAY-2000.
XX
PF   04-NOV-1999;   99WO-US26039.
XX
PR   05-NOV-1998;   98US-0107243.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z,  Schading RL;
XX
DR   WPI; 2000-376566/32.
DR   N-PSDB; AAD00668.
XX
PT   Application of a hypersensitive response elicitor protein to plants to
PT   impart stress resistance
XX
PS   Disclosure; Page 7-8; 84pp; English.
XX
CC   The patent discloses a method to impart stress resistance to plants by
CC   applying a hypersensitive response elicitor in a non-infectious form to
CC   a plant or seed. The present sequence is a hypersensitive
CC   response elicitor protein from Erwinia amylovora. It is
CC   is used to impart stress resistance to plants.
XX
SQ   Sequence   403 AA;

Alignment Scores:
Pred. No.:              1.61e-161       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
DB:                     21              Gaps:           0

SEQ12-SEQ4 (1-1390) x AAY71093 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCC 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy    235 GGCGGAAATAACGGCTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGCGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy    355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGCCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTCCTGAAGATGTTCAGCGAGATAATGCAAGCCTGTTTGGTGATGGGCAAGATGCCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GCAGTCACTGATGCGCTGTCGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGCCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GCGGCACTGGCAGGTGGTCAGGCGCGTAATGCTGGCACGGGTCTTGACGGTTCCTCCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeuLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGACCGGCCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 CTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy    1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGCGATCGGGCGATGGCGAAGGAAATC 1074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy    1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGCCCG 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy    1135 GGTCAGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGACGCAAGCCAGATGAC 1194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy    1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy    1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy    1315 CTGGGTATTGATGCCATGATGGCCGGTGATCCCATTAACAATATGGCACTTGGCAAGCTG 1374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy    1375 GGCGCGGCT 1383
           |||||||||
Db    401 GlyAlaAla 403

RESULT 6
AAY84854
ID   AAY84854 standard; Protein; 403 AA.
XX
AC   AAY84854;
XX
DT   08-AUG-2000  (first entry)
XX
DE   A hypersensitive response elicitor protein.
XX
KW   Hypersensitive response; insect control; disease resistance;
KW   hypersensitive response elicitor; plant growth; vegetable; crop;
KW   ornamental plant.
XX
OS   Erwinia amylovora.
XX
FH   Key             Location/Qualifiers
FT   Peptide         169..403
```

```
FT                   /note= "claimed under claim 5"
FT   Peptide         210..403
FT                   /note= "claimed under claim 5"
FT   Peptide         267..403
FT                   /note= "claimed under claim 5"
FT   Peptide         343..403
FT                   /note= "claimed under claim 5"
FT   Peptide         105..179
FT                   /note= "claimed under claim 7"
FT   Peptide         121..150
FT                   /note= "claimed under claim 7"
FT   Peptide         137..166
FT                   /note= "claimed under claim 7"
FT   Peptide         137..156
FT                   /note= "claimed under claim 7"XX
XX
PN   WO200020452-A2.
XX
PD   13-APR-2000.
XX
PF   05-OCT-1999;    99WO-US23181.
XX
PR   05-OCT-1998;    98US-0103050.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z,  Fan H,  Niggemeyer JL;
XX
DR   WPI; 2000-303745/26.
DR   N-PSDB; AAA14938.
XX
PT   Hypersensitive response elicitor pOlypeptides useful for imparting
PT   enhanced growth, disease resistance and insect resistance to plants,
PT   especially vegetables and ornamental flowers
XX
PS   Claim 4; Page 8-10; 100pp; English.
XX
CC   The present sequence represents  a hypersensitive response elicitor
CC   polypeptide. The specification describes hypersensitive response
CC   elicitor polypeptide fragments, which do not elicit a hypersensitive
CC   response. Instead, the proteins impart disease resistance to plants,
CC   enhance plant growth, and/or control insects. The polypeptide
CC   fragments may be used to these properties to plants. The plants which
CC   may be treated in this way include vegetables, crops and ornamental
CC   plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,
CC   peanut, corn, potato, sweet potato, bean, pea, chicory, lettuce,
CC   endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,
CC   broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,
CC   carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,
CC   citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
CC   tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,
CC   petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.
XX
SQ   Sequence   403 AA;

Alignment Scores:
Pred. No.:          1.61e-161       Length:         403
Score:              2079.00         Matches:        403
Percent Similarity: 100.00%         Conservative:   0
Best Local Similarity: 100.00%      Mismatches:     0
Query Match:        82.11%          Indels:         0
DB:                 21              Gaps:           0

SEQ12-SEQ4 (1-1390) x AAY84854 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy    235 GGCGGAAATAACGGGTTGCTCGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy    355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCYGATGGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTCAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGCGCCTGATGGCTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGCTACCGGTATCGGTATGAAAGCGGGCATTCAGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgGlyMetAlaGlyLysGluIle 300

Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGCAATGACACCAGCCAGTATGCAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
```

```
                    |||||||||
Db      401 GlyAlaAla 403

RESULT 7
AAE06710
ID   AAE06710 standard; Protein; 403 AA.
XX
AC   AAE06710;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Erwinia amylovora hypersensitive response elicitor protein.
XX
KW   Hypersensitive response elicitor; oomycete; transgenic plant; infection;
KW   gene therapy; crOp loss; antifungal.
XX
OS   Erwinia amylovora.
XX
PN   WO200155347-A1.
XX
PD   02-AUG-2001.
XX
PF   26-JAN-2001; 2001WO-US02579.
XX
PR   26-JAN-2000; 2000US-0178565.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Beer SV,  Bauer DW;
XX
DR   WPI; 2001-488791/53.
DR   N-PSDB; AAD12806.
XX
PT   New chimeric gene, useful for controlling plant-pathogenic fungi and
PT   producing oomycete-resistant transgenic plants, comprises first DNA
PT   encoding hypersensitive response elicitor, promoter and regulatory
PT   region -
XX
PS   Claim 9; Page 13-14; 72pp; English.
XX
CC   The invention relates to a chimeric gene that includes a first DNA
CC   molecule encoding a hypersensitive response elicitor protein or
CC   polypeptide, promoter operably linked 5' to the first DNA molecule
CC   to induce transcription of the first DNA molecule in response to
CC   activation of the promoter by an oomycete and a 3' regulatory region
CC   operably linked to the first DNA molecule. The invention also relates
CC   to a transgenic plant resistant to disease resulting from oomycete
CC   infection, the transgenic plant including the chimeric gene, wherein
CC   the promoter induces transcription of the first DNA molecule in
CC   response to infection of the plant by an oomycete. The chimeric gene
CC   is used in gene therapy. The chimeric gene is useful as an effective
CC   and safe means of controlling plant-pathogenic fungi, particularly
CC   oomycetes, which are responsible for major crop loss and is also useful
CC   for producing transgenic plants of the invention. The present sequence
CC   is Erwinia amylovora hypersensitive response elicitor protein.
XX
SQ   Sequence   403 AA;

Alignment Scores:
Pred. No.:              1.61e-161       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
DB:                     22              Gaps:           0

SEQ12-SEQ4 (1-1390) x AAE06710 (1-403)

Qy      175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy      235 GGCGGAAATAACGGGTTGCTGGCTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy      295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy      355 ACCGGCATGATGATGATGACGATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ThrGlyMetMetMetMetMetSerMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy      415 GGCGGCTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       81 GlyGlyGlyLeuGlyAsnGlyLeuGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy      475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy      535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy      595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy      655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy      715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GlnGlySerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy      775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy      835 GGGGGACTGGGAGGTGGTCAGGGCCGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      221 GlyGlyLeuGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy      895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy      955 GTGGGTACCGGTATCGGTATGAAAGCCGGCCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy     1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      281 ArgHisSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy     1075 GGTCAGTTCATGGACCACTATCCTGAGGTGTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy     1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy     1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy     1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy     1315 CTGGGTATTGATGCCATGATGGCCCGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

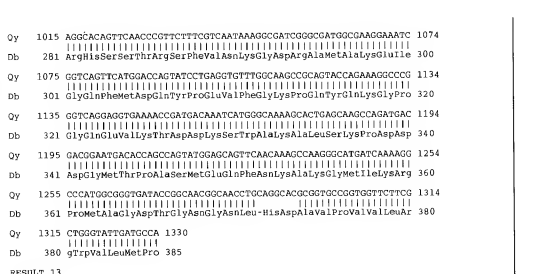
```
Db     381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy    1375 GGCGCGGCT 1383
           |||||||||
Db     401 GlyAlaAla 403

RESULT 8
ABB09225
ID   ABB09225 standard; Protein; 403 AA.
XX
AC   ABB09225;
XX
DT   08-JUL-2002  (first entry)
XX
DE   Erwinia amylovora hypersensitive response elicitor SEQ ID NO:3.
XX
KW   Hypersensitive response elicitor; fruit; vegetable; plant; desiccation;
KW   postharvest disease.
XX
OS   Erwinia amylovora.
XX
PN   WO200180639-A2.
XX
PD   01-NOV-2001.
XX
PF   17-APR-2001; 2001WO-US12468.
XX
PR   19-APR-2000; 2000US-198359P.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z,  Qiu D,  Remick D;
XX
DR   WPI; 2002-041357/05.
DR   N-PSDB; ABL51710.
XX
PT   Inhibiting post harvest disease (caused by Penicillium, Botrytis,
PT   Phytophthora, or Erwinia) or desiccation and enhancing the longevity in
PT   a fruits or vegetables, using hypersensitive response elicitor proteins
PT   or nucleic acids -
XX
PS   Example; Page 9-10; 72pp; English.
XX
CC   The present invention describes methods for inhibiting post harvest
CC   disease or desiccation and enhancing the longevity in a fruits or
CC   vegetables, using hypersensitive response elicitor proteins or
CC   polypeptides or nucleic acids (I) derived from pathogens (e.g. Erwinia
CC   amylovora, E. stewartii, E. chrysanthemi, E. carotovora, Xanthomonas,
CC   Pseudomonas syringae, P. solanacearum, Phytophthora, and Clavibacter).
CC   (I) has bactericidal activity, and can be used in gene therapy. The
CC   method can be used for inhibiting post harvest disease (caused by
CC   Penicillium, Botrytis, Phytophthora, or Erwinia) or desiccation and
CC   enhancing the longevity in a fruits or vegetables. The method enables
CC   growers, warehouse packers, shippers and suppliers to process, handle
CC   and store fruit and vegetables with reduced loses caused by post harvest
CC   disease and desiccation, therefore reducing costs to the consumer and
CC   improving quality. The present sequence represents a hypersensitive
CC   response elicitor protein given in the exemplification of the present
CC   invention.
XX
SQ   Sequence   403 AA;

Alignment Scores:
Pred. No.:              1.61e-161       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
DB:                     23              Gaps:           0

SEQ12-SEQ4 (1-1390) x ABB09225 (1-403)

Qy     175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCC 234
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy     235 GGCGGAAATAACGGGTTGCTGGTACCAGTCCCCAGAATGCTGGGTTGGGTGGCAATTCT 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy     295 GCACTGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy     355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGCTGATGGGCCGTGGCTTA 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy     415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy     475 GCGCTGAACGATATGTTAGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy     535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy     595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy     655 CTGCTGAACATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy     715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy     775 GGACTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy     835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGTG 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy     895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy     955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy     1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGCCGATGGCGAAGGAAATC 1074
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 ArgHisSerGlnThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy     1075 GGTCAGTTCATGGACCAGTATCCTGAGGTCTTTGGCAACCGCACTACCAGAAAGGCCCG 1134
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy     1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy     1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy     1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db    401 GlyAlaAla 403

RESULT 9
AAE18295
ID   AAE18295 standard; Protein; 403 AA.
XX
AC   AAE18295;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Erwinia amylovora hypersensitive response elicitor (HRE) #1.
XX
KW   Hypersensitive response elicitor; HRE; transgenic plant; plant growth;
KW   stress tolerance; disease tolerance; modified flower colour;
KW   insect resistance; herbicide resistance; male sterility.
XX
OS   Erwinia amylovora.
XX
PN   WO200195724-A2.
XX
PD   20-DEC-2001.
XX
PF   13-JUN-2001; 2001WO-US18955.
XX
PR   15-JUN-2000; 2000US-211585P.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z,  Derocher J;
XX
DR   WPI; 2002-130707/17.
DR   N-PSDB; AAD29124.
XX
PT   Improving effectiveness of transgenic plants by topical application of
PT   a hypersensitive response elicitor protein to the transgenic plant or
PT   by incorporating into the plant a transgene encoding the protein  -
XX
PS   Disclosure; Page 10-11; 86pp; English.
XX
CC   The invention relates to methods of improving the effectiveness of
CC   transgenic plants which involves either topical application of a
CC   hypersensitive response elicitor (HRE) protein to the transgenic plant
CC   or incorporating into the transgenic plant a transgene encoding HRE.
CC   HRE sequence is used for improving the effectiveness of transgenic
CC   plants by maximising the benefit of transgenic traits associated with
CC   a deleterious effect on growth, stress tolerance, disease or insect
CC   resistance, enhanced growth, herbicide resistance, male sterility,
CC   modified flower colour and biochemically modified plant product in
CC   the transgenic plants or overcoming the deleterious effects. The
CC   present sequence is Erwinia amylovora HRE protein.
XX
SQ   Sequence  403 AA;

Alignment Scores:
Pred. No.:              1.61e-161       Length:         403
Score:                  2079.00         Matches:        403
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            82.11%          Indels:         0
DB:                     23              Gaps:           0

SEQ12-SEQ4 (1-1390) x AAE18295 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy    355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACAGCGTGGCCTCGAAAGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCGAGATAATGCAAGCCTGTTTGGTGATGGGCAAGATGCGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGCGATCGGGCGATGGCGAAGGAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerThrArgSerProValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGC 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380
```

```
Qy   1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db    401 GlyAlaAla 403

RESULT 10
AAE16447
ID   AAE16447 standard; Protein; 403 AA.
XX
AC   AAE16447;
XX
DT   09-APR-2002  (first entry)
XX
DE   E. amylovora hypersensitive response elicitor protein #1.
XX
KW   Hypersensitive response elicitor protein; plant growth; fruit coloration;
KW   disease resistance; stress resistance; phytotoxin; insect infection;
KW   plant maturation.
XX
OS   Erwinia amylovora.
XX
FH   Key             Location/Qualifiers
FT   Domain          32..74
FT                   /label= Hypersensitive_response_eliciting_domain_1
FT   Region          32..57
FT                   /label= Acidic_unit
FT   Region          57..74
FT                   /label= Alpha_helix
FT   Domain          130..180
FT                   /label= Hypersensitive_response_eliciting_domain_2
FT   Region          130..157
FT                   /label= Acidic_unit
FT   Region          157..180
FT                   /label= Alpha_helix
XX
PN   WO200198501-A2.
XX
PD   27-DEC-2001.
XX
PF   12-JUN-2001; 2001WO-US18820.
XX
PR   16-JUN-2000; 2000US-212211P.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Fan H,  Wei Z;
XX
DR   WPI; 2002-122282/16.
DR   N-PSDB; AAD27015.
XX
PT   New hypersensitive response elicitor proteins comprising spaced apart
PT   domains having an acidic portion linked to an alpha-helix, useful for
PT   imparting disease or stress resistance, controlling insects or
PT   enhancing plant growth -
XX
PS   Disclosure; Page 9-10; 99pp; English.
XX
CC   The patent discloses hypersensitive response elicitor proteins and
CC   nucleotides encoding such proteins. Hypersensitive response elicitor
CC   proteins comprise an isolated pair or more of spaced apart domains,
CC   each comprising an acidic portion linked to an alpha-helix and capable
CC   of eliciting a hypersensitive response in plants. Sequences of the
CC   invention are used to impart disease resistance to plants, to enhance
CC   plant growth, to control insects and/or to impart stress resistance
CC   to plants which includes resistance to environmental stresses such as
CC   climate, air pollution, chemical and nutritional stress. The method of
CC   imparting disease resistance has the potential for treating previously
CC   untreatable diseases, treating diseases systemically and avoiding the
CC   use of infectious agents or environmentally harmful materials. Hyper-
CC   sensitive response elicitor sequences are used to enhance plant growth
CC   which encompasses greater yield, increased in quantity of seeds produced,
CC   percentage of seeds germinated, plant size and biomass, bigger fruits,
CC   earlier fruit coloration and plant maturation. They are also used for
CC   insect control which encompasses preventing direct insect damage to
CC   plant by feeding injury, interfering with insect larval feeding on the
CC   plants, preventing insects from colonising host plants and releasing
CC   phytotoxins. Sequences of the invention also prevent subsequent disease
CC   damage to plants resulting from insect infection. The present sequence
CC   is Erwinia amylovora hypersensitive response elicitor protein.
XX
SQ   Sequence   403 AA;

Alignment Scores:
Pred. No.:              1.61e-161       Length:          403
Score:                  2079.00         Matches:         403
Percent Similarity:     100.00%         Conservative:    0
Best Local Similarity:  100.00%         Mismatches:      0
Query Match:            82.11%          Indels:          0
Db:                     23              Gaps:            0

SEQ12-SEQ4 (1-1390) x AAE16447 (1-403)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGCCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy    355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGCGGTTCCCTGAACACGCTGGGCTCCAAAGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGCGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGCGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy    655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy    955 GTGGGTACCGGTATCGCTATCGAAAGCGGCCATTCAGCCGCTCAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTCAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGluGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGlnAlaArgGlyAlaGlyGlySerSer 380

Qy   1315 CTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAGCTG 1374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 LeuGlyIleAspAlaMetMetAlaGlyAspAlaIleAsnAsnMetAlaLeuGlyLysLeu 400

Qy   1375 GGCGCGGCT 1383
          |||||||||
Db    401 GlyAlaAla 403

RESULT 11
AAW06598
ID   AAW06598 standard; Protein; 385 AA.
XX
AC   AAW06598;
XX
DT   30-MAR-1997  (first entry)
XX
DE   Hypersensitive response elicitor protein.
XX
KW   Hypersensitive response; elicitor; Erwinia amylovora; plant;
KW   disease-resistance; Escherichia coli; infiltration; virus;
KW   bacterium; fungus; pathogen; biological control agent.
XX
OS   Erwinia amylovora.
XX
PN   WO9639802-A1.
XX
PD   19-DEC-1996.
XX
PF   05-JUN-1996;    96WO-US08819.
XX
PR   07-JUN-1995;    95US-0475775.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Beer SV,  Wei Z;
XX
DR   WPI; 1997-051614/05.
DR   N-PSDB; AAT49314.
XX
PT   Imparting pathogen resistance to plants - with hypersensitive
PT   response elicitor polypeptide or protein
XX
PS   Claim 7; Page 46-47; 69pp; English.
XX
CC   This sequence represents a hypersensitive response elicitor from
CC   Erwinia amylovora, with a mol.wt. of 37,000.  The elicitor has a pI
CC   of 4.3, thermostability at 100 deg C for at least 10 min, and
CC   contains no cysteine.  The elicitor may be used in a new method for
CC   imparting pathogen resistance to plants, by application of the
CC   elicitor in a non-infectious form to plant cells, by spraying,
CC   injection, leaf abrasion, or plant infection with recombinant
CC   bacteria (non-infectious to the host plant, e.g. Escherichia coli)
CC   expressing the elicitor as a biological control agent, to alllow
CC   recombinant protein infiltration into the plant.  The method
CC   confers virus, bacterium or fungus disease-resistance on crops and
CC   ornamental plants.
XX
SQ   Sequence   385 AA;

Alignment Scores:
Pred. No.:              6.33e-154       Length:          385
Score:                  1986.00         Matches:         385
Percent Similarity:     99.74%          Conservative:    0
Best Local Similarity:  99.74%          Mismatches:      0
Query Match:            78.44%          Indels:          1
DB:                     18              Gaps:            0

SEQ12-SEQ4 (1-1390) x AAW06598 (1-385)

Qy    175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTCGG 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy    235 GGCGGAAATAACGGGTTGCTGGGTACCAGTCCCCAGAATGCTGGGTTGGGTGGCAATTCT 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy    295 GCACTGGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 AlaLeuGlyLeuGlyGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy    355 ACCGGCATGATGATGATGACCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ThrGlyMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy    415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCCAAC 474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy    475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGCGGCAACAAT 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy    535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAA 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy    595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerProGlnMetGlnGlnGln 160

Qy    655 CTGCTGAACATGTTCAGCGAGATAATGCAAACCCTGTTTGCTGATGGGCAAGATGGCACC 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy    715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGCGAGCAGAACGCCTATAAAAAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGlnGlnAsnAlaTyrLysLys 200

Qy    775 GGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGCCAAC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy    835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy    895 GCCGGCAAAGGGCTGCAAAACCTGACCGGCGGCCGGTGGACTACCAGCAGTTAGCTAACGCC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GlyGlyLysGlyLeuGlnAsnLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260
```

```
Qy    955 GTGGGTACCGGTATCGGTATGAAAGCCGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280

Qy   1015 AGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGTAATC 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ArgHisSerSerThrArgSerPheValAsnLysGlyAspArgAlaMetAlaLysGluIle 300

Qy   1075 GGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGCCCG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GlyGlnPheMetAspGlnTyrProGluValPheGlyLysProGlnTyrGlnLysGlyPro 320

Qy   1135 GGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGATGAC 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GlyGlnGluValLysThrAspAspLysSerTrpAlaLysAlaGluSerLysProAspAsp 340

Qy   1195 GACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAAAGG 1254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 AspGlyMetThrProAlaSerMetGlnGlnPheAsnLysAlaLysGlyMetIleLysArg 360

Qy   1255 CCCATGCGCGGCTGATACCGGCAACCGCAACCTGCAGGCACGCGGTGCCGGTGGTTCTTCG 1314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ProMetAlaGlyAspThrGlyAsnGlyAsnLeuGln-HisAlaValProValLeuAr 380

Qy   1315 CTGGGTATTGATGCCA 1330
          ||||||||||||||||
Db    380 gTrpValLeuMetPro 385

RESULT 12
AAR45751
ID   AAR45751 standard; Protein; 385 AA.
XX
AC   AAR45751;
XX
DT   11-JUL-1994  (first entry)
XX
DE   Erwinia amylovora harpin.
XX
KW   Harpin; hypersensitive response elicitor; HR-elicitor; fire blight;
KW   Rosaceae; apple; pear; phytopathogenic bacteria; defence reaction;
KW   hypersensitive reaction and pathogenicity; hrpN; gene cluster; ds.
XX
OS   Erwinia amylovora.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 372
FT                   /note= "corresponds to CAG codon"
FT   Misc-difference 373
FT                   /note= "corresponds to CAC codon"
XX
PN   WO9401546-A.
XX
PD   20-JAN-1994.
XX
PF   30-JUN-1993;   93WO-US06243.
XX
PR   01-JUL-1992;   92US-0907935.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Bauer DW,  Beer SV,  Collmer A,  He S,  Laby R,  Wei Z;
XX
DR   WPI; 1994-035054/04.
DR   N-PSDB; AAQ55751.
XX
PT   Hypersensitive response elicitor protein derived from Erwinia
PT   amylovora - and DNA encoding it, useful for developing harpin
PT   inhibitors to prevent e.g. fire blight of fruit
XX
PS   Claim 6; Page 27-28; 47pp; English.
XX
CC   The hrpN gene was isolated from E.amylovora using a 48-fold
CC   degenerate oligonucleotide probe corresponding to amino acids 9-15
CC   at the N-terminus of harpin. The 44kD protein encoded by the hrpN
CC   gene is a hypersensitive response elicitor protein. The harpin is
CC   thought to be an archetype for HR elicitors from phytopathogenic
CC   bacteria.
XX
SQ   Sequence   385 AA;

Alignment Scores:
Pred. No.:                  5.78e-152       Length:          385
Score:                      1962.00         Matches:         381
Percent Similarity:         98.96%          Conservative:    1
Best Local Similarity:      98.70%          Mismatches:      3
Query Match:                77.49%          Indels:          1
DB:                         15              Gaps:            0

SEQ12-SEQ4 (1-1390) x AAR45751 (1-385)

Qy     175 ATGAGTCTGAATACAAGTGGGCTGGGAGCGTCAACGATGCAAATTTCTATCGGCGGTGCG 234
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetSerLeuAsnThrSerGlyLeuGlyAlaSerThrMetGlnIleSerIleGlyGlyAla 20

Qy     235 GGCGCGAAATAACGGGTTGCTGGCTACCAGTCGCCAGAATGCTGGGTTGGGTGGCAATTCT 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 GlyGlyAsnAsnGlyLeuLeuGlyThrSerArgGlnAsnAlaGlyLeuGlyGlyAsnSer 40

Qy     295 GCACTGGGCTGGGCGGCGGTAATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTC 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 AlaLeuGlyLeuGlyGlyAsnGlnAsnAspThrValAsnGlnLeuAlaGlyLeuLeu 60

Qy     355 ACCGGCATGATGATGATGATGAGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTA 414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ThrGlyMetMetMetMetMetSerMetMetGlyGlyGlyGlyLeuMetGlyGlyGlyLeu 80

Qy     415 GGCGGTGGCTTAGGTAATGGCTTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAAC 474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 GlyGlyGlyLeuGlyAsnGlyLeuGlyGlySerGlyGlyLeuGlyGluGlyLeuSerAsn 100

Qy     475 GCGCTGAACGATATGTTAGGCGGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAAT 534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 AlaLeuAsnAspMetLeuGlyGlySerLeuAsnThrLeuGlySerLysGlyGlyAsnAsn 120

Qy     535 ACCACTTCAACAACAAATTCCCCGCTGGACCAGCGCTGGGTATTAACTCAACGTCCCAA 594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ThrThrSerThrThrAsnSerProLeuAspGlnAlaLeuGlyIleAsnSerThrSerGln 140

Qy     595 AACGACGATTCCACCTCCGGCACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAG 654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AsnAspAspSerThrSerGlyThrAspSerThrSerAspSerSerAspProMetGlnGln 160

Qy     655 CTGCTGAAGATGTTCAGCGAGATAATGCAAAGCCTGTTTGGTGATGGGCAAGATGGCACC 714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     161 LeuLeuLysMetPheSerGluIleMetGlnSerLeuPheGlyAspGlyGlnAspGlyThr 180

Qy     715 CAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAAAAA 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GlnGlySerSerSerGlyGlyLysGlnProThrGluGlyGluGlnAsnAlaTyrLysLys 200

Qy     775 GGAGTCACTGATGGCGTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGCAAC 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 GlyValThrAspAlaLeuSerGlyLeuMetGlyAsnGlyLeuSerGlnLeuLeuGlyAsn 220

Qy     835 GGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCGCTG 894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GlyGlyLeuGlyGlyGlyGlnGlyGlyAsnAlaGlyThrGlyLeuAspGlySerSerLeu 240

Qy     895 GGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCACTTAGGTAACGCC 954
           ||||||||||||:::  ||||||||||||||||||||||||||||||||||||||||||||
Db     241 GlyGlyLysGlyLeuArgGlyLeuSerGlyProValAspTyrGlnGlnLeuGlyAsnAla 260

Qy     955 GTGGGTACCGGTATCGGTATGAAAGCGGCATTCAGGCGCTGAATGATATCGGTACGCAC 1014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ValGlyThrGlyIleGlyMetLysAlaGlyIleGlnAlaLeuAsnAspIleGlyThrHis 280
```



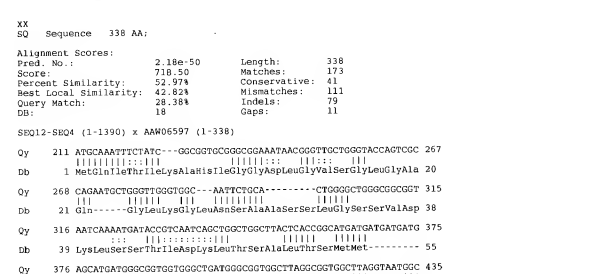
RESULT 13
AAW06597
ID AAW06597 standard; Protein; 338 AA.
XX
AC AAW06597;
XX
DT 30-MAR-1997 (first entry)
XX
DE Hypersensitive response elicitor protein.
XX
KW Hypersensitive response; elicitor; Erwinia chrysanthemi; plant;
KW disease-resistance; Escherichia coli; infiltration; virus;
KW bacterium; fungus; pathogen; biological control agent.
XX
OS Erwinia chrysanthemi.
XX
PN WO9639802-A1.
XX
PD 19-DEC-1996.
XX
PF 05-JUN-1996; 96WO-US08819.
XX
PR 07-JUN-1995; 95US-0475775.
XX
(CORR) CORNELL RES FOUND INC.
XX
PI Beer SV, Wei Z;
XX
DR WPI; 1997-051614/05.
DR N-PSDB; AAT49313.
XX
PT Imparting pathogen resistance to plants - with hypersensitive
PT response elicitor polypeptide or protein
XX
PS Claim 4; Page 44; 69pp; English.
XX
CC This sequence represents a hypersensitive response elicitor from
CC Erwinia chrysanthemi, with a mol.wt. of 34,000. The elicitor is
CC thermostable, has a glycine content of over 16%, and has no cysteine.
CC The elicitor may be used in a new method for imparting pathogen
CC resistance to plants, by application of the elicitor in a non-
CC infectious form to plant cells, by spraying, injection, leaf
CC abrasion, or plant infection with recombinant bacteria (non-
CC infectious to the host plant, e.g. Escherichia coli) expressing the
CC elicitor as a biological control agent, to allow recombinant
CC protein infiltration into the plant. The method confers virus,
CC bacterium or fungus disease-resistance on crops and ornamental
CC plants.

XX
SQ Sequence 338 AA;

Alignment Scores:

Pred. No.:	2.18e-50	Length:	338
Score:	718.50	Matches:	173
Percent Similarity:	52.97%	Conservative:	41
Best Local Similarity:	42.82%	Mismatches:	111
Query Match:	28.38%	Indels:	79
DB:	18	Gaps:	11

SEQ12-SEQ4 (1-1390) x AAW06597 (1-338)



```
Qy   1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
          |||||||||||||||||||||||||||||||:::|||||||||||||:::|||||||||
Db    235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy   1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGCAAAAGCACTGAGCAAGCCAGAT 1191
                           ||||||||||||||||||||||||||||||||||||||||||||
Db    255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy   1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
          ||||||||||||||    |||||||||::::::|||    :::|||    |||||||||||
Db    275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy   1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
                :::||||||||||||||||||    ||||||      ||||||||||||||:::
Db    295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy   1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
          |||||||||||||||||    :::    ||||||    |||    ||||||:::||||||||
Db    315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy   1372 CTGGGCGCGGCT 1383
          |||      |||
Db    335 LeuAlaAsnAla 338
```

RESULT 14
AAW87638

```
ID   AAW87638 standard; Protein; 338 AA.
XX
AC   AAW87638;
XX
DT   09-MAR-1999  (first entry)
XX
DE   A hypersensitive response elicitor protein.
XX
KW   Hypersensitive response elicitor protein; hairpin protein;
KW   disease resistance; seed quality; insect control; corn borer;
KW   Lepidoptera larvae; transgenic plant.
XX
OS   Erwinia chrysanthemi.
XX
PN   WO9854214-A2.
XX
PD   03-DEC-1998.
XX
PF   28-MAY-1998;   98WO-US10874.
XX
PR   30-MAY-1997;   97US-0048109.
XX
PA   (CORR ) CORNELL RES FOUND INC.
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Beer SV,  Laby RJ,  Wei Z;
XX
DR   WPI; 1999-070210/06.
DR   N-PSDB; AAV83988.
XX
PT   New fragments of an Erwinia hypersensitive response elicitor protein
PT   and related DNA - used to impart disease resistance to plants, to
PT   increase their growth and to control insects
XX
PS   Disclosure; Page 7-8; 94pp; English.
XX
CC   The present sequence represents a hypersensitive response elicitor
CC   protein (also called hairpin protein) that is able to elicit a
CC   hypersensitive response in plants. The specification also describes
CC   hypersensitive response elicitors from other pathogenic organisms.
CC   The protein, in non-infectious form, is applied to plants to impart
CC   disease resistance (to a wide range of viral, bacterial and fungal
CC   pathogens), to improve growth (yield, quantity and quality of seeds,
CC   to provide earlier germination etc.) and to control insects (e.g. corn
CC   borers, Lepidoptera larvae etc.) The same results are provided by
CC   transgenic plants expressing the protein.
XX
SQ   Sequence   338 AA;

Alignment Scores:
Pred. No.:              2.18e-50     Length:        338
Score:                  718.50       Matches:       173
Percent Similarity:     52.97%       Conservative:  41
Best Local Similarity:  42.82%       Mismatches:    111
Query Match:            28.38%       Indels:        79
DB:                     20           Gaps:          11

SEQ12-SEQ4 (1-1390) x AAW87638 (1-338)

Qy    211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
          |||||||||:::|||          ||||||:::   |||:::   |||
Db      1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy    268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
          |||      ||||||   |||   |||||||||         ||||||
Db     21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy    316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                :::   |||:::::::::|||           ||||||   ||||||
Db     39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy    376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                            ||||||   |||      |||
Db     56 ------------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy    436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
          ||||||   |||         |||   |||:::||||||   |||               |||
Db     64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy    496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                            |||::::::|||   :::|||      |||
Db     83 Asn--------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy    556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                  ||||||
Db     97 ------------------------------------------------------SerGly 98

Qy    616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
             |||:::   |||                             ||||||||   :::
Db     99 GlyAspAlaLeuSer-----------------------------LysMetPheAspLys 108

Qy    676 ATAATGCAAAGCCTGTTTGGT------------------------GATGGGCAAGATGGC 711
             :::   |||   |||                           :::        :::
Db    109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy    712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                |||                   |||   |||:::|||      ||||||:::
Db    129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy    772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
             ||||||   :::|||||||||   ::::::|||||||||||   |||    :::
Db    149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy    832 AACGGGGGACTGGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                 :::|||           |||
Db    168 ---------------------------------------SerGlyPheSerGlnProSer 174

Qy    892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
          ||||||      |||||||||   |||||||||         :::   |||||||||||
Db    175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy    952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
          |||:::|||   |||:::|||         |||   :::   ||||||:::::::::   |||
Db    195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy   1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
          |||         :::   |||   |||||||:::|||   ||||||   |||||||||||
Db    215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy   1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
```

```
              ||||||||||||||||||||||||||||||||:::||||||||||||:::|||||||||
Db      235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy     1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                              ||||||||||||||||||||||||||||||||||||||||||
Db      255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy     1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
            ||||||||||||||    |||||||||::::::|||    :::|||   ||||||||||||
Db      275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294

Qy     1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
                 :::||||||||||||||||||    ||||||      ||||||||||||||||:::
Db      295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy     1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
            ||||||||||||||||||    :::    ||||||    |||   ||||||:::|||||||||
Db      315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy     1372 CTGGGCGCGGCT 1383
            |||      |||
Db      335 LeuAlaAsnAla 338

RESULT 15
AAW82407
ID   AAW82407 standard; Protein; 338 AA.
XX
AC   AAW82407;
XX
DT   23-FEB-1999  (first entry)
XX
DE   E. chrysanthemi HrpN-ECH protein.
XX
KW   Hypersensitive response elicitor protein; hrpN-Ech; pathogen resistance;
KW   plant; transformation; pathogen-inducible promoter.
XX
OS   Erwinia chrysanthemi.
XX
PN   US5850015-A.
XX
PD   15-DEC-1998.
XX
PF   07-JUN-1995;   95US-0484358.
XX
PR   07-JUN-1995;   95US-0484358.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Bauer D,  Collmer A;
XX
DR   WPI; 1999-069852/06.
DR   N-PSDB; AAV73494, AAV73507.
XX
PT   DNA encoding Erwinia chrysanthemi hypersensitive response protein
PT   hrpN - useful for imparting pathogen resistance to plants
XX
PS   Claim 2; Column 29-30; 27pp; English.
XX
CC   This sequence represents a novel Erwinia chrysanthemi protein, hrpN-Ech,
CC   that elicits a hypersensitive response in plants. The encoding DNA can be
CC   used for imparting pathogen resistance to plants, by transforming a
CC   plant with a vector containing the DNA and a pathogen-inducible promoter.
XX
SQ   Sequence   338 AA;

Alignment Scores:
Pred. No.:              2.18e-50        Length:          338
Score:                  718.50          Matches:         173
Percent Similarity:     52.97%          Conservative:    41
Best Local Similarity:  42.82%          Mismatches:      111
Query Match:            28.38%          Indels:          79
DB:                     20              Gaps:            11
```

```
SEQ12-SEQ4 (1-1390) x AAW82407 (1-338)

Qy      211 ATGCAAATTTCTATC---GGCGGTGCGGGCGGAAATAACGGGTTGCTGGGTACCAGTCGC 267
            ||||||||:::|||          ||||||:::    |||:::    |||
Db        1 MetGlnIleThrIleLysAlaHisIleGlyGlyAspLeuGlyValSerGlyLeuGlyAla 20

Qy      268 CAGAATGCTGGGTTGGGTGGC---AATTCTGCA---------CTGGGGCTGGGCGGCGGT 315
            |||      ||||||   |||   |||||||||           ||||||
Db       21 Gln------GlyLeuLysGlyLeuAsnSerAlaAlaSerSerLeuGlySerSerValAsp 38

Qy      316 AATCAAAATGATACCGTCAATCAGCTGGCTGGCTTACTCACCGGCATGATGATGATGATG 375
                :::   |||:::::::::|||          ||||||   ||||||
Db       39 LysLeuSerSerThrIleAspLysLeuThrSerAlaLeuThrSerMetMet--------- 55

Qy      376 AGCATGATGGGCGGTGGTGGGCTGATGGGCGGTGGCTTAGGCGGTGGCTTAGGTAATGGC 435
                                                  ||||||    |||       |||
Db       56 ----------------------------------PheGlyGlyAlaLeuAlaGlnGly 63

Qy      436 TTGGGTGGCTCAGGTGGCCTGGGCGAAGGACTGTCGAACGCGCTGAACGATATGTTAGGC 495
            ||||||   |||            |||   |||:::||||||   |||          |||
Db       64 LeuGlyAlaSer---SerLysGlyLeuGlyMetSerAsnGlnLeuGlyGlnSerPheGly 82

Qy      496 GGTTCGCTGAACACGCTGGGCTCGAAAGGCGGCAACAATACCACTTCAACAACAAATTCC 555
                              |||:::::::|||   :::|||      |||
Db       83 Asn--------------GlyAlaGlnGlyAlaSerAsnLeuLeuSerValProLys--- 96

Qy      556 CCGCTGGACCAGGCGCTGGGTATTAACTCAACGTCCCAAAACGACGATTCCACCTCCGGC 615
                                                                  ||||||
Db       97 ------------------------------------------------------SerGly 98

Qy      616 ACAGATTCCACCTCAGACTCCAGCGACCCGATGCAGCAGCTGCTGAAGATGTTCAGCGAG 675
               |||:::    |||                            |||||||||    :::
Db       99 GlyAspAlaLeuSer----------------------------LysMetPheAspLys 108

Qy      676 ATAATGCAAAGCCTGTTTGGT----------------------GATGGGCAAGATGGC 711
               :::      |||   |||                         :::      :::
Db      109 AlaLeuAspAspLeuLeuGlyHisAspThrValThrLysLeuThrAsnGlnSerAsnGln 128

Qy      712 ACCCAGGGCAGTTCCTCTGGGGGCAAGCAGCCGACCGAAGGCGAGCAGAACGCCTATAAA 771
                        |||            |||   |||:::|||      ||||||:::
Db      129 LeuAlaAsnSerMetLeuAsnAlaSerGlnMetThrGlnGlyAsnMetAsnAlaPheGly 148

Qy      772 AAAGGAGTCACTGATGCGCTGTCGGGCCTGATGGGTAATGGTCTGAGCCAGCTCCTTGGC 831
               ||||||   :::|||||||||   ::::::||||||||||||   |||    :::
Db      149 SerGlyValAsnAsnAlaLeuSerSerIleLeuGlyAsnGlyLeuGlyGlnSerMet--- 167

Qy      832 AACGGGGGACTGGAGGTGGTCAGGGCGGTAATGCTGGCACGGGTCTTGACGGTTCGTCG 891
                                                   :::|||            |||
Db      168 ------------------------------------SerGlyPheSerGlnProSer 174

Qy      892 CTGGGCGGCAAAGGGCTGCAAAACCTGAGCGGGCCGGTGGACTACCAGCAGTTAGGTAAC 951
            ||||||      |||||||||   |||||||||            :::   |||||||||||
Db      175 LeuGlyAlaGlyGlyLeuGlnGlyLeuSerGlyAlaGlyAlaPheAsnGlnLeuGlyAsn 194

Qy      952 GCCGTGGGTACCGGTATCGGTATGAAAGCGGGCATTCAGGCGCTGAATGATATCGGTACG 1011
            |||:::|||   |||:::|||         |||   :::   ||||||::::::::::   |||
Db      195 AlaIleGlyMetGlyValGlyGlnAsnAlaAlaLeuSerAlaLeuSerAsnValSerThr 214

Qy     1012 CACAGGCACAGTTCAACCCGTTCTTTCGTCAATAAAGGCGATCGGGCGATGGCGAAGGAA 1071
            |||         :::   |||   ||||||:::|||   ||||||   |||||||||||
Db      215 HisValAspGlyAsnAsnArgHisPheValAspLysGluAspArgGlyMetAlaLysGlu 234

Qy     1072 ATCGGTCAGTTCATGGACCAGTATCCTGAGGTGTTTGGCAAGCCGCAGTACCAGAAAGGC 1131
            ||||||||||||||||||||||||||||||||:::||||||||||||:::|||||||||
Db      235 IleGlyGlnPheMetAspGlnTyrProGluIlePheGlyLysProGluTyrGlnLysAsp 254

Qy     1132 CCGGGTCAGGAGGTGAAAACCGATGACAAATCATGGGCAAAAGCACTGAGCAAGCCAGAT 1191
                              ||||||||||||||||||||||||||||||||||||||||||
Db      255 GlyTrpSerSerProLysThrAspAspLysSerTrpAlaLysAlaLeuSerLysProAsp 274

Qy     1192 GACGACGGAATGACACCAGCCAGTATGGAGCAGTTCAACAAAGCCAAGGGCATGATCAAA 1251
            ||||||||||||||    |||||||||::::::|||    :::|||   ||||||||||||
Db      275 AspAspGlyMetThrGlyAlaSerMetAspLysPheArgGlnAlaMetGlyMetIleLys 294
```

```
Qy    1252 AGGCCCATGGCGGGTGATACCGGCAACGGCAACCTGCAGGCACGCGGTGCCGGTGGTTCT 1311
               :::|||||||||||||||||||    ||||||      ||||||||||||||:::
Db     295 SerAlaValAlaGlyAspThrGlyAsnThrAsnLeuAsnLeuArgGlyAlaGlyGlyAla 314

Qy    1312 TCGCTGGGTATTGATGCCATGATGGCCGGTGATGCCATTAACAATATGGCACTTGGCAAG 1371
           |||||||||||||||||  :::   ||||||   |||   |||||||:::|||||||||
Db     315 SerLeuGlyIleAspAlaAlaValValGlyAspLysIleAlaAsnMetSerLeuGlyLys 334

Qy    1372 CTGGGCGCGGCT 1383
           |||     |||
Db     335 LeuAlaAsnAla 338

Search completed: January 14, 2003, 17:08:33
Job time : 54.5 secs
```